Reproductive function in female childhood cancer survivors

Overbeek, A.

2018

Reproductive function in female childhood cancer survivors

Annelies Overbeek

516278-L-bw-Overbeek
Processed on: 19-1-2018

ISBN: 978-94-028-0920-6

The research described in this thesis was performed at the Division of Pediatric Oncology, Department of Pediatrics and the Department of Obstetrics and Gynecology, VU University Medical Center, Amsterdam, the Netherlands.

Financial support by the Dutch Cancer Society (grant no. VU 2006-3622) and by Foundation Children Cancer Free (grant no. 2008-20) is gratefully acknowledged.
Additional support was provided by Philips and Durex.
Financial support for printing of this thesis was also kindly provided by Stichting Research Kindergeneeskunde, SWOG, VUmc, Stichting Fertiliteitsfonds, Goodlife Pharma, Origio Benelux, Titus Healthcare, Chipsoft and Teva Nederland.

Reproductive function in female childhood cancer survivors
Thesis, VU University Medical Center Amsterdam, the Netherlands
Layout design: Rens Dommerholt, Persoonlijk Proefschrift,
www.persoonlijkproefschrift.nl
Layout based on photo by Jeffrey van der Steen, Prime Fotografie,
www.primefotografie.nl
Printed by: Ipskamp printing, www.ipskampprinting.nl

VRIJE UNIVERSITEIT

Reproductive function in female childhood cancer survivors

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op vrijdag 16 maart 2018 om 09.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Annelies Overbeek
geboren te Alkmaar

promotoren:	prof.dr. G.J.L. Kaspers prof.dr. F.E. van Leeuwen prof.dr. C.B. Lambalk
copromotor:	dr. E. van Dulmen-den Broeder

With special gratitude to the review committee
Professor D.D.M. Braat, MD, PhD, Radboud University Medical Center, Nijmegen
Professor M.J.C. Eijkemans, MD, PhD, University Medical Center Utrecht, Utrecht
Professor B.C.J.M Fauser, MD, PhD, University Medical Center Utrecht, Utrecht
Professor M. Goddijn, MD, PhD, Academic Medical Center, Amsterdam
Professor C.J.M. de Groot, MD, PhD, VU University Medical Center, Amsterdam
Professor R. Pieters, MD, PhD, Princess Máxima Center for Pediatric Oncology, Utrecht
Professor W.H. Wallace, MD, PhD, Royal Hospital for Sick Children, Edinburgh

TABLE OF CONTENTS

1	General introduction	9
2	Chemotherapy-related late adverse effects on ovarian function in female survivors of childhood cancer and cancer in the reproductive age: a systematic review *Cancer Treat Rev. 2017 Feb;53:10-24.*	25
3	A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges *BMC Cancer. 2012 Aug 23;12(1):363.*	55
4	Fertility studies in female childhood cancer survivors: selecting appropriate comparison groups *Reprod Biomed Online. 2014 Sep;29(3):352-61.*	73
5	Validity of self-reported data on pregnancies for childhood cancer survivors: a comparison with data from a nationwide population-based registry *Hum Reprod. 2013 Mar;28(3):819-27.*	89
6	Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics *J Med Internet Res. 2011 Sep 29;13(3):e76.*	105
7	Intra-cycle fluctuations of anti-Müllerian hormone in normal women with a regular cycle: a re-analysis *Reprod Biomed Online. 2012 Jun;24(6):664-9.*	121
8	Between-method agreement and inter-rater reliability of antral follicle counts using conventional and three dimensional ultrasound in childhood cancer survivors *In preparation.*	131
9	Comparison of ovarian function markers in users of hormonal contraceptives during the hormone-free interval and subsequent natural early follicular phases *Hum Reprod. 2010 Jun;25(6):1520-7.*	147

10	Long-term effects of childhood cancer treatment on ovarian function markers: results of a Dutch nationwide study *Submitted.*	**163**
11	Practice, Attitude and Knowledge of Dutch pediatric oncologists regarding female fertility *Neth J Med. 2014 Jun;72(5):264-70.*	**213**
12	General discussion, clinical implications and recommendations for future research	**227**
13	Nederlandse samenvatting	**241**
&	Appendices	**247**
R	References	**263**

1

General introduction

OVARIAN PHYSIOLOGY

In order to understand the effect of chemotherapy and radiotherapy on ovarian function, it is important to briefly explain ovarian physiology and the menstrual cycle. All women are born with a finite pool of oocytes. This pool is said to be non-replenishable. Initially there are 6-7 million primordial follicles present at the fourth month of fetal development, whereas only 400,000-600,000 primordial follicles remain at menarche, and finally, at the stage of menopause, the number of remaining follicles has dropped to below 1000. This age-dependent demise of oocytes is illustrated in Figure 1, depicting the total population of non-growing follicles, calculated from eight histological studies in which numbers of non-growing follicles were assessed from conception to menopause.

The rate of non-growing follicle recruitment increases from birth to 14 years and then declines with age until menopause. The model shows that 81% of the variance in total follicle pool is due to age [1].

Age of natural menopause in the developed world ranges between 40 and 58 years, with an approximate mean of 51.4 years. Ethnicity as well as lifestyle-, reproductive and genetic factors (smoking, parity, age at menarche) are associated with age at menopause [2]. Premature menopause is diagnosed when amenorrhea is present before the age of 40 years with hypersecretion of FSH. There are many terms used for this condition, including premature ovarian failure, primary ovarian failure, primary ovarian insufficiency, premature ovarian insufficiency and hypergonadotropic hypogonadism.

Although the actual cessation of menses occurs, on average, in the fifth decade of life, fecundity already starts to decrease some twenty years earlier, with subfertility setting in from age 31 years and beyond, and sterility around the age of forty [3]. It is known that with the advancement of age, the quality of the oocytes gradually deteriorates. The stages in the natural transition to menopause are shown in Figure 2.

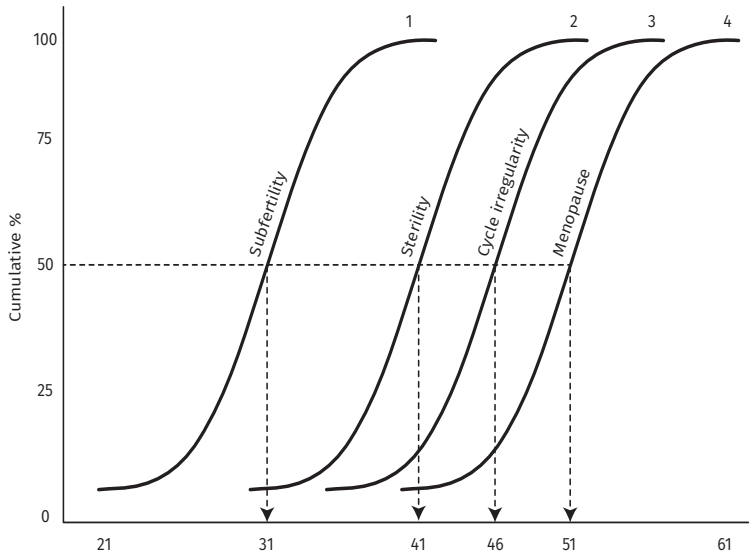


Figure 2 The distributions of age at reproductive events, Lambalk et al, Maturitas 2009 [3].

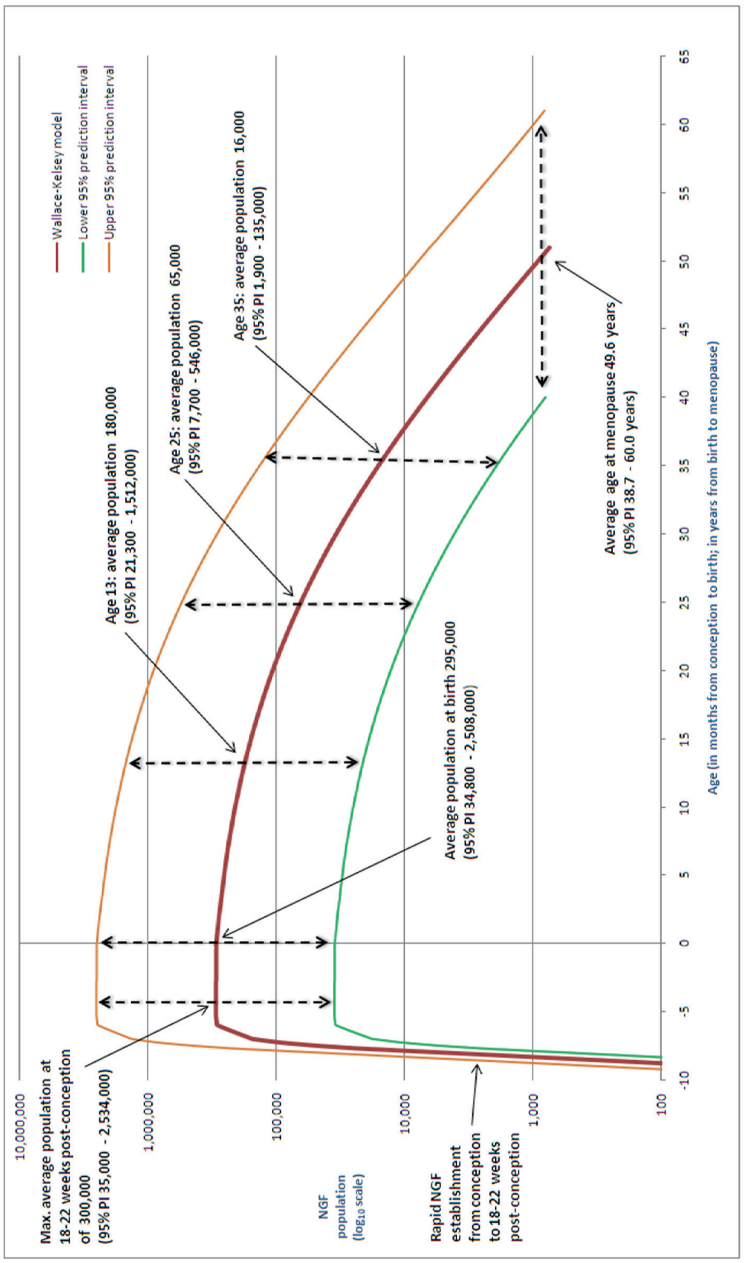


Figure 1 Values for non-growing follicle populations at illustrative ages, together with the corresponding 95% prediction intervals. Wallace et al, PLoS one, 2010 [1].

At the basis of the gradual decline of the oocyte pool from puberty to menopause is the occurrence of a monthly menstrual cycle during which the number of oocytes decrease due to ovulation and atresia. During the reproductive phase of a woman's life, a cohort of primordial follicles is continuously recruited to grow, in response to signals that are independent of gonadotropins. One of these signals has been identified as anti-Müllerian hormone (AMH). Subsequent enlargement of the oocyte, in response to AMH, leads to proliferation of the granulosa cells and the formation of a pre-antral follicle. This follicle, in turn, develops follicle-stimulating hormone-receptors in the granulosa cells over the next 3–6 months, after which the follicle forms a fluid-filled space called an antrum (antral follicle). Autocrine factors are required for an adequate progression through these successive stages of follicle development [4]. The antral follicle becomes dependent on follicle-stimulating hormone (FSH) for further development [5]. AMH limits the FSH sensitivity, thereby further regulating follicular recruitment [6]. A drop in sex steroid and inhibin A levels at the end of the previous cycle will result in an increase in gonadotropin releasing hormone (GnRH) and subsequent increased gonadotropin levels from the pituitary [7] that further sustain follicle growth and follicular steroidogenesis. As the follicles become larger in the early mid-follicular phase, they produce more estrogen and inhibin, which will gradually decrease the serum levels of FSH, due to negative feedback. Smaller follicles, with fewer FSH receptors, do no longer receive sufficient FSH to continue growing and will undergo atresia. Therefore, generally only one follicle reaches the stage of ovulation in the natural cycle. Rapidly increasing levels of estradiol produced by the mature preovulatory follicle, precede the midcycle LH and FSH surge that will initiate ovulation. The corpus luteum is formed by angiogenesis and luteinisation of the granulosa and theca cells. This leads to an alteration of the steroidogenic pathway, which results in the production of progesterone. Inhibin A, which is mainly produced by the corpus luteum, might increase the production of progesterone [4]. The corpus luteum retains the ability to produce estrogen. When no pregnancy occurs, LH levels become too low to sustain the corpus luteum, initiating regression and thereby leading to a fall in progesterone and estrogen and the onset of menses. FSH levels rise with withdrawal of estrogens negative feedback and the next cohort of follicles begins to develop (Figure 3) [7].

This monthly process inevitably causes the follicle pool to deplete, inaugurating the menopausal transition. Lower amounts of follicles result in lower inhibin B production and lower AMH concentrations. Subsequently FSH concentrations increase. Both decreased AMH and increased FSH concentrations lead to a higher percentage of growing follicles relative to the resting pool. This phenomenon results in an acceleration of follicle depletion. Because of this increased follicle recruitment, a normal menstrual pattern will be maintained, despite ovarian aging. The first clinical sign of ovarian aging in women is shortening of the cycle. Higher basal FSH concentrations cause earlier follicle recruitment. This starts in the final stages of the previous cycle before menstruation resulting in a virtually shorter follicular phase of the following cycle and higher estradiol concentrations at the start of the cycle. Approximately one year prior to menopause, 60–70% of the menstrual cycles become anovulatory or have a prolonged follicular phase. In this period, estrogen concentrations fluctuate strongly from undetectable to up to many times normal.

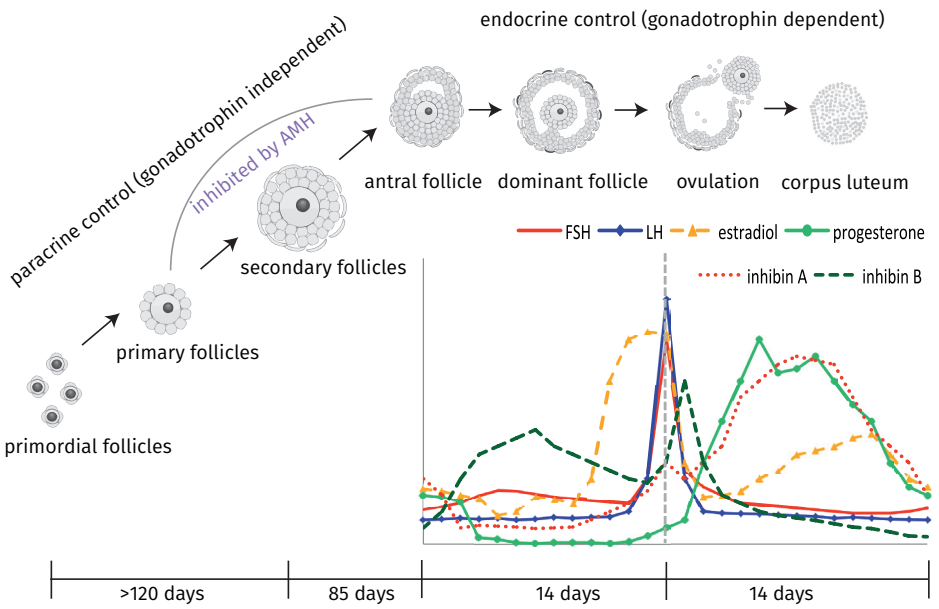


Figure 3 Stages of follicle recruitment and selection in the human ovary including a graph of hormone levels during the menstrual cycle. Adapted from McGee and Hsueh (2000) and Verhoeven et al. (2016) [8 9].

These irregular fluctuations are insufficient to trigger an LH surge, and persisting follicles occur. When the resting pool of oocytes consists of 1,000 follicles or less menopause occurs, since the ovary is not able to maintain the cyclic hormonal production for regulation of the normal menstrual cycle [8].
Cycle irregularities (either shortening or elongation of the cycle) are thus important clinical markers of impending menopause.

HOW CAN OVARIAN FUNCTION BE MEASURED?

There are many different ways to assess ovarian reserve and ovarian function. Careful history taking, with special attention to family history, cycle shortening and occurrence of vasomotor symptoms, can provide tools to aid in the prediction of age at menopause.
When ovarian function is assessed, tests typically concentrate on the functional reserve of active antral follicles. They do not measure the resting pool of primordial follicles [10]. The most widely used methods focus either on measurements of hormones in the blood, produced by the ovary (AMH, inhibin B) or the pituitary (FSH), or on ultrasonographic measurements of the ovaries (antral follicle count, ovarian volume). Inhibin A, LH, estradiol and progesterone are not considered reliable markers for assessing ovarian reserve [3].
When the total number of follicles decreases due to impending menopause, FSH levels rise and menstrual cycles will start to shorten. Serum FSH levels above 10

U/L are generally acknowledged as indicative of an impaired ovarian function, if measured in the early follicular phase at least twice. Inhibin B is secreted from granulosa cells in FSH-dependent growing follicles and levels are correlated with the number of developing antral follicles seen on ultrasonography during the early follicular phase [11]. It has been shown that inhibin B levels drop in parallel with the number of ovarian antral follicles [12]. No specific concentration of inhibin B has been shown to be diagnostically discriminatory [13]. AMH plays a role in inhibiting follicular recruitment and FSH-dependent growth, and it may regulate the selection of pre-antral and small antral follicles [14-16]. Its cut-off value for low ovarian reserve in literature is rather arbitrary, ranging between 0.1 and 1 ng/ml [17-19]. Also, the laboratory method used to measure serum AMH influences its value [20]. Of the endocrinological markers, AMH seems superior in that it decreases gradually with age, as opposed to FSH and inhibin B that only change at the end of the reproductive life span. Thus, it is the first marker to detect a decrease in ovarian function in the sequence of events leading to menopause. Furthermore, some studies [21 22], but not all [23-28], show that AMH sampling does not seem to require specific timing and its reproducibility between cycles is good.
Antral follicle counts are measured by transvaginal ultrasound and this technique concentrates on the small antral follicles (2-10 mm) that can be visualized in the early follicular phase. Antral follicle counts have been found to correspond well with histological analysis (which is the gold standard) of the follicle pool and show a gradual decline with age [29]. The fact that antral follicle counts as well as AMH gradually decrease with age, allows for predictions of ovarian reserve relatively early in life. However, all these markers are parameters of the dynamic profile of the ovaries, and they have not been shown to accurately predict fecundability [30], or menopause [31].

CHILDHOOD CANCER AND LATE EFFECTS

In 2016, in the Netherlands, 586 children below the age of 18 years were diagnosed with childhood cancer [32]. The most common types of childhood cancer are hematological malignancies, followed by solid tumors and tumors of the brain and central nervous system.
Over the past 40 years there have been major improvements in the treatment of childhood cancer, tremendously increasing survival [33-35]. As a result, there is a rapidly growing population of survivors of childhood cancer. In 2014, 6,168 5-years survivors were registrated in the Netherlands [32]. As a consequence of the increased survival, more childhood cancer survivors live to be confronted with late effects of the cancer treatment, which may not become apparent until decades later. Approximately 3 out of 4 survivors experience one or more adverse long-term health events [36 37].
This increase in the occurrence of late effects on many organ systems has prompted pediatric oncologists to continuously evaluate and change the treatment protocols for childhood cancer.

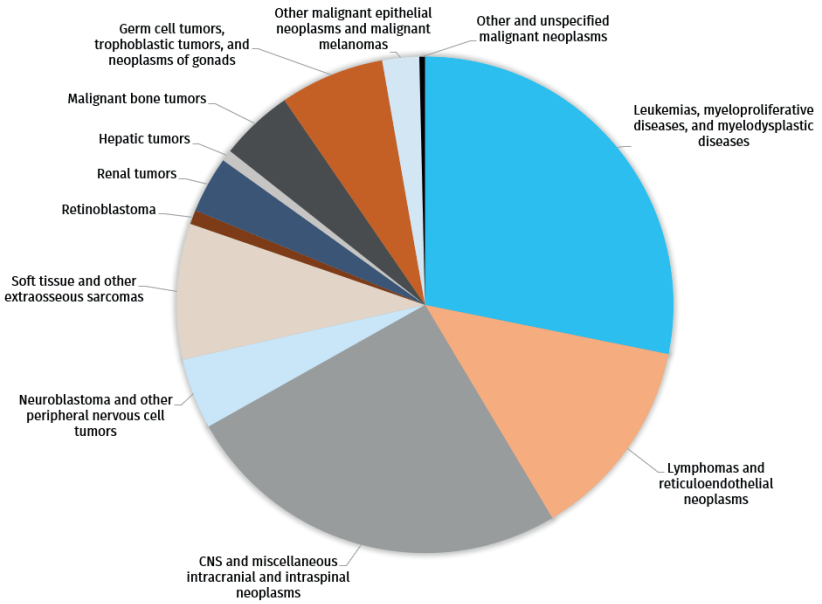


Figure 4 Types of childhood cancer in the Netherlands in 2014. Skion Basisregistratie 2017 [32].

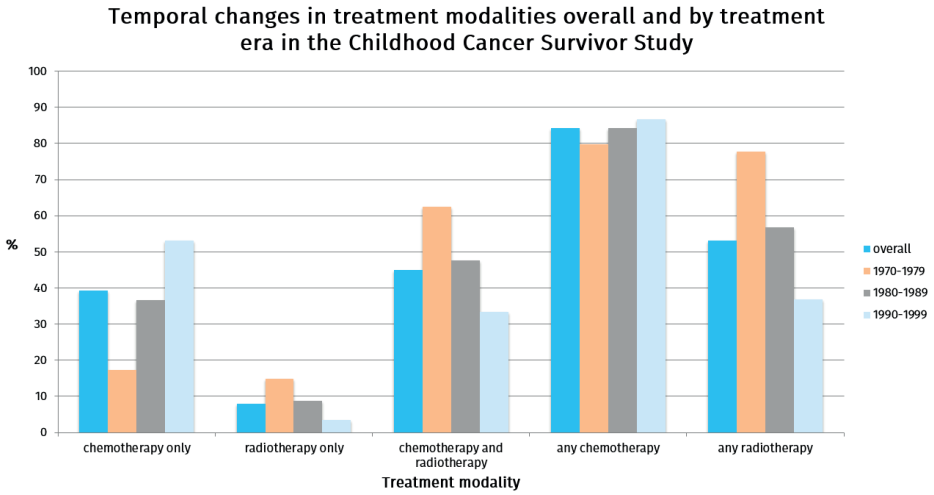


Figure 5 Temporal changes in treatment modalities, overall and by treatment era, described in the cohort of the Childhood Cancer Survivor Study. Adapted from Table 1 from Turcotte et al. JAMA 2017 [38].

In a recent report of the Childhood Cancer Survivor Study, temporal changes in treatment in 5-year survivors were evaluated [38]. Complete treatment data were available for 83% of the cohort. Between 1970-1999, there were substantial changes in therapies. Radiation therapy decreased from 77% of survivors treated in the 1970s to 54% in the 1980s and 33% in the 1990s. Median radiation treatment dose decreased from 30Gy (interquartile range, 24-44) in the 1970s to 26 Gy (interquartile range, 18-52) in the 1990s. Although the proportion of children treated with alkylating agents and anthracyclines increased over time, median doses decreased. The proportion of children treated with epipodophyllotoxins and platinum agents also increased over the three decades; however, whereas the median cumulative dose of platinum increased with each treatment decade, the median cumulative dose of epipodophyllotoxins increased substantially in the 1980s and then decreased in the 1990s. The most important changes are depicted in Figure 5.

In females, a compromised reproductive system is an important and frequently encountered late effect of treatment, which has high impact on quality of life [39 40]. Cancer treatment may damage the hypothalamic-pituitary-ovarian-uterine axis, leading to delayed or arrested puberty, sub- or infertility and adverse pregnancy outcomes [39 41-43]. However, it may also reduce the fertile life span and induce premature menopause, since therapy may deplete or may accelerate the decline of the non-renewable pool of primordial follicles in the ovary [17 44]. The issue of premature menopause has only gained attention in the last two decades since it is only now that significant numbers of female childhood cancer survivors (CCSs) are reaching their forties. The available literature suggests that approximately 6% of CCS will experience acute ovarian failure (loss of ovarian function within 5 yr of diagnosis), with those diagnosed with Hodgkin lymphoma, treated at older age and with high doses of ovarian or abdominal radiotherapy being at highest risk [45]. In one of the earlier studies (1992) it was found that over 40% of childhood cancer survivors had experienced menopause at the age of 31 [46]. A more recent report on childhood cancer survivors found that age at menopause was earlier but rarely premature, with a median age of 44 years (range 18-55 years) [47]. In the St. Judes lifetime cohort the prevalence of premature ovarian insufficiency (POI) was 10.9%. Independent risk factors for POI included ovarian radiotherapy at any dose and a cyclophosphamide equivalent dose (CED) of ≥8,000 mg/m^2. Surprisingly, patients with a BMI ≥30 kg/m^2 at the time of the assessment were less likely to have a diagnosis of POI [48]. Nevertheless, even in the most recent studies, the median age of the CCS is around 32 years, so a large proportion of the cohort has not reached the ages at which premature menopause can be established or excluded. The prevalence of premature menopause is thus expected to increase with a longer follow-up.

The differences in age of menopause in various studies over time may in part be explained by the ever-changing treatment protocols for the different types of childhood cancer. As more knowledge on late effects becomes available, treatment protocols are further adapted to reduce the prevalence and severity of late effects without compromising efficacy.

Table 1 Different classes of chemotherapy and their method of action. Generic names as well as commonly used brand names are stated

Class	Method of action	Type	Generic name (Brand name)
Alkylating agents	Crosslink guanine nucleobases in DNA	Nitrogen mustards	Mechlorethamine (Mustargen®), Ifosfamide (Ifex®), Melphalan (Alkeran®)/L-PAM/ phenylalanine mustard, Chlorambucil (Leukeran®), Cyclophosphamide (Cytoxan®, Neosar®)
		Alkyl sulfonates	Busulfan (Myleran®), Treosulfan (Ovastat®)
		Ethylene imines	Thiotepa (Thioplex®, Girostan®, Thiofozil®, Tifosyl®), Altretamine (Hexalen®)
		Nitrosureas	Carmustine (Becenum®, BCCNU, Carmibris, Gliadel), Lomustine (CCNU), Semustine (methyl-CCNU), Streptozocin (Zanosar®), Fotemustine (Muphoran®), Nimustine (ACNU, Nidran), Ranimustine (Cymer, Cymerin)
		Triazenes	Dacarbazine (DTIC-Dome), temozolomide (Temodar)
Antimetabolites	Inhibit DNA synthesis	Pyrimidine compounds	5-flurouracil (5-FU), arabinosylcytosine, capecitabine, gemcitabine, decitabine
		Purine compounds	Fludarabine, 6-mercaptopurine (6-MP), thioguanine (6-TG), cladribine (Leukostatin), Pentostatin (Nipent)
		Folate antagonists	Methotrexate, Pemetrexed
Platinum agents	Crosslinking of DNA, thereby inhibiting DNA repair of synthesis		Cisplatin, carboplatin, oxaloplatin, nedaplatin
Mitotic inhibitors	Inhibit function of microtubules		Vinblastine, Vinorelbine (Navelbine), Vincristine (Oncovin), Vindesine, Etoposide, Teniposide,
Antitumor antibiotics	Induce DNA lesions, inhibit topo-isomerase, among other effects	Anthracyclines	Daunorubicin, doxorubicin (Adriamycin), epirubicin, idarubicin, valrubicin, bleomycin, mitomycin, plicamycin, dactonimycin

CANCER TREATMENT AND OVARIAN FUNCTION

Adverse effects of radiotherapy and chemotherapy on ovarian function

It has been demonstrated that abdominal and pelvic radiotherapy has deleterious effects on ovarian function. The degree of harm is related to the localisation, total radiation dose, fractionation schedule and age at time of treatment [49]. Cranial irradiation (> 35–40 Gy) can impair the hypothalamic pituitary function, resulting in hypogonadism through GnRH or FSH/LH deficiency. Radiation to the pelvis may, in addition to causing severe damage to the ovarian follicle pool, induce uterine fibrosis, which may complicate future pregnancies.

In contrast to these adverse late effects of radiotherapy on ovarian function, the exact mechanism by which specific chemotherapeutics do harm in females remains unclear. Histopathology of the ovary has shown that chemotherapy destroys the primordial follicles and can lead to ovarian atrophy in animal models. Other mechanisms causing damage to the ovary include injury to blood vessels and focal ovarian cortical fibrosis [50]. Recently, the 'burn-out' hypothesis was postulated, explaining the loss of follicles after chemotherapy. This hypothesis is based on research with genetically modified mice that do not produce anti-Mullerian hormone (AMH-null). The mice show increased activation of primordial follicles, leading to increased atresia of large follicles, and consequently accelerating the depletion of the total follicle pool. Treatment with chemotherapy destroys growing follicles and these follicles are therefore prevented from producing AMH. AMH inhibits primordial follicle recruitment, so a drop in AMH levels might lead to an increase in large follicles and an accelerated decrease in primordial follicles [51]. The damage that is induced by chemotherapy is drug-, dose- and age- dependent. The following chemotherapeutic agents have been associated with gonadotoxicity: alkylating agents (including procarbazine), antimetabolites, platinum agents, taxanes, and possibly vinca-alkaloids [52]. Table 1 describes for each class of gonadotoxic chemotherapeutic agent the method of action as well as the most frequently prescribed chemotherapy agents.

Ovarian function in cancer survivors

The effects of radio- and chemotherapy on ovarian function have been evaluated in several studies, some by using self-reported data, others (in general smaller studies) by using different ovarian function tests. The most important study conducted in a large cohort of childhood cancer survivors (CCSs) concluded that risk factors for nonsurgical premature menopause (based on self report) included attained age, exposure to increasing doses of radiation to the ovaries, increasing alkylating agent score (based on number of alkylating agents and cumulative dose), and a diagnosis of Hodgkin lymphoma. For survivors who were treated with alkylating agents plus abdominopelvic radiation, the cumulative incidence of nonsurgical premature menopause approached 30% [53]. Radiotherapy consisting of more than 10 Gy to the abdomen has been shown to cause ovarian failure. In

a multivariable logistic regression model, increasing doses of ovarian irradiation, exposure to procarbazine, and exposure to cyclophosphamide at ages 13-20 yr were independent risk factors for acute ovarian failure [45]. Clinically assessed ovarian dysfunction was primarily described in survivors of Hodgkin lymphoma or childhood leukemia, and was assessed by pubertal development and levels of traditional endocrine markers of gonadal function (FSH, LH, estradiol). These, mostly small, studies showed that ovarian dysfunction was mainly observed after treatment with abdominal-pelvic radiotherapy, total body irradiation, and/or alkylating agents (cyclophosphamide and procarbazine in particular). However, as the rise in FSH occurs late in the sequence of events associated with ovarian aging, the increase in FSH is of limited use as a clinical marker for reproductive counselling. Ultrasound-based antral follicle counts (AFC) and AMH are considered to offer a higher clinical value for assessing ovarian function. Again, procarbazine, alkylating chemotherapy in general and ovarian irradiation have been shown to influence AMH and AFC [17 19 54-57]. A recent study showed that cyclophosphamide and ifosfamide did not show a significant reduction in ovarian function, as measured by antral follicle counts, AMH, FSH and ovarian surface, whereas previous studies have shown a negative effect of cyclophosphamide on ovarian function [57]. A complete overview of the literature with regard to the effect of chemotherapy on ovarian function is summarized in Chapter 2.

Within the International Late Effects of Childhood Cancer Guideline Harmonization Group (IGHG), which aims to establish a common vision and integrated strategy for the surveillance of chronic health problems in childhood, adolescent, and young adult cancer survivors, a recommendation has been made for the surveillance for POI in female childhood cancer survivors. This recommendation is based on all the available evidence on this topic up until 2014. In Table 2 the level of evidence of the association of treatment factors and the risk of premature ovarian insufficiency in survivors of childhood cancer and young adult cancer is summarized [58].

Gaps in knowledge

Table 2 depicts clearly that for most chemotherapeutic agents, no solid evidence on premature ovarian insufficiency is available to assess its gonadotoxicity. Specific knowledge gaps were formulated in the recent IHGH guideline for POI surveillance in childhood cancer survivors: (1) there is no clear evidence to indicate which type of alkylating agent chemotherapy increases the risk of POI, and (2), there is no information regarding a safe threshold dose for chemotherapy as well as for radiotherapy [58].

Most studies derived from the Childhood Cancer Survivor Study, as well as other studies that chose amenorrhea as an endpoint, have used self-reported questionnaires to assess menopause, which may bias results. Most reports clinically assessing reproductive function in childhood survivors have focused on small study populations, often describing 10-150 patients, with various types of childhood cancer. This heterogeneity in types of tumors as well as the use of multimodal therapy make lack of power one of the largest culprits in assessing late effects

in childhood cancers, and has restricted the possibility to establish robust dose-effect relationships. In addition, many studies have not used a full panel of ovarian reserve tests. Because AFC and AMH show a gradual decline with age, these may be valuable markers to assess ovarian reserve in CCSs at a young age. However, more data are needed to evaluate the diagnostic and prognostic value of these markers with regard to chance of pregnancy or age at menopause.
To date, most studies have relied on predictive models derived from cross-sectional data. As the inter-individual variance in markers is substantial, longitudinal assessment of ovarian function will make a predictive model for ovarian reserve and prediction of age at menopause more valid.

Table 2 Evidence levels for the risk of premature ovarian insufficiency in relation to treatment in female survivors of childhood and young adult cancer. Derived from van Dorp et al., Journal of Clinical Oncology (2016) [58]

Risk of premature ovarian insufficiency	Level of evidence*
Increased risk after alkylating agents vs. no alkylating agents	Level A
Increased risk after higher alkylating agent dose vs. lower dose	Level A
Increased risk after cyclophosphamide vs. no cyclophosphamide	Level C
Increased risk after higher cyclophosphamide dose vs. lower dose	No studies
Increased risk after procarbazine vs. no procarbazine	Level C
Increased risk after higher procarbazine dose vs. lower dose	No studies
Risk after multiple alkylating agents and other chemotherapeutic agents vs. single alkylating agents	No studies
Risk after other alkylating agents (busulfan, chlorambucil, mechloramine, ifosfamide, melphalan, thiotepa, carmustine, lomustine)	No studies
Risk after platinum agents (carboplatin and cisplatin)	No studies
Increased risk after radiotherapy to which ovaries are potentially exposed vs. no radiotherapy	Level A
Increased risk after higher dose of radiotherapy to which ovaries are potentially exposed vs. lower dose	Level A
Increased risk after radiotherapy to which ovaries are potentially exposed and alkylating agents vs. either treatment in the same dose alone	Level C
Increased risk after treatment at older age vs. younger age	Level B
Risk after unilateral oophorectomy	No studies

**Levels of evidence are based on the criteria of the American Heart Association. Level A: High-quality evidence from more than 1 RCTs; meta-analyses of high-quality RCTs; one or more RCTs corroborated by high-quality registry studies. Level B: Moderate-quality evidence from 1 or more RCTs; meta-analyses of moderate-quality RCTs; moderate-quality evidence‡ from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies; meta-analyses of such studies. Level C: Randomized or nonrandomized observational or registry studies with limitations of design or execution; meta-analyses of such studies; physiological or mechanistic studies in human subjects; consensus of expert opinion based on clinical experience.*

AIM OF THIS THESIS

Based on the studies available, we concluded that a future study should include larger numbers of long-term CCSs and adequate control groups, and focus on a prolonged and more complete follow-up of these CCSs (to prevent selection bias). Furthermore, an extensive set of clinical ovarian function markers should be used to assess ovarian function. Since some of the ovarian function markers are influenced by cycle phase or exogenous hormone use, these factors should be taken into account. Treatment data should be as detailed as possible, including radiation dosimetry. Data should subsequently be linked to actual fertility and occurrence of premature menopause, since the few larger studies that suggested a relationship between treatment and risk of early menopause obtained their data from questionnaires, without any additional clinical ascertainment.
Therefore, in 2008 we started the DCOG LATER-VEVO study (VEVO is a Dutch acronym for 'Vruchtbaarheid, Eicelvoorraad en Vervroegde Overgang') from which the first results are presented in this thesis. This retrospective cohort study, performed within the nationwide collaborative group called 'Dutch Childhood Oncology Group – Long-term Effects after Childhood Cancer (DCOG LATER)', examines the effects of childhood cancer treatment on the reproductive system of female CCSs, treated between 1963 and 2002 at one of the seven Dutch pediatric oncology - and stem cell transplant centers using an adequate control group and a full panel of ovarian function and ovarian reserve tests. In addition, the effects of different types and doses of treatment, including radiation sites and type of chemotherapeutic regimen, were determined.

OUTLINE OF THIS THESIS

In **Chapter 2**, we review the literature and perform a qualitative synthesis on the chemotherapy-related adverse effects on ovarian function in female CCS. In **Chapters 3** and **4**, we describe the design of the DCOG LATER-VEVO study, as well as the pitfalls we have encountered while conducting the study. In addition, we describe our experiences regarding the recruitment of subjects for two comparison groups for the DCOG LATER-VEVO study.
In order to optimally interpret the results of the DCOG LATER-VEVO study, we performed several validation studies with regard to the methods we used (questionnaire, endocrinological markers, transvaginal ultrasound measurements). In **Chapters 5** and **6**, we focus on the accuracy of pregnancy outcomes reported in the DCOG-LATER VEVO questionnaire by CCSs and sibling controls and assess whether sending a mixed invitation (paper-based and web-based) rather than an invitation for a web-based questionnaire only resulted in different response and participation rates. **Chapter 7** describes the intra-cycle variation of AMH in healthy women, which emphasizes the difficulties that arise when searching for reliable ovarian reserve tests. The reproducibility and agreement of 2D and 3D transvaginal ultrasound measurements of the antral follicle counts are presented in **Chapter 8**. Many women participating in the DCOG LATER-VEVO study (CCSs as well as controls) were using

hormonal contraceptives and it was not known whether reproductive function could be accurately assessed under such conditions. In **Chapter 9**, we therefore evaluate whether values of FSH, LH, estradiol, AMH, inhibin B, antral follicle count (AFC) and ovarian volume (OV) determined on day 7 of the hormone-free interval are similar to values measured on days 2–5 of two subsequent natural menstrual cycles.
In **Chapter 10**, we discuss the main results of the DCOG LATER-VEVO study, regarding the effect of cancer therapy on ovarian function. In addition, we report on the validity and agreement of the different ovarian function markers in CCS.
While conducting the DCOG LATER-VEVO study, it became clear that knowledge about fertility issues and fertility preservation among Dutch pediatric oncologists was limited. Therefore, we conducted the PAK-study to assess the current practice, attitudes, and knowledge of Dutch pediatric oncologists involved in oncological care regarding fertility and fertility preservation options in female childhood cancer patients. The results of this study are presented in **Chapter 11**.
Finally, **Chapter 12** presents a general discussion of our findings as well as clinical implications and recommendations for future research. **Chapter 13** concludes this thesis with a Dutch summary.

516278-L-bw-Overbeek
Processed on: 19-1-2018

2

Chemotherapy-related late adverse effects on ovarian function in female survivors of childhood cancer and cancer in the reproductive age: a systematic review

A. Overbeek
M.H. van den Berg
F.E. van Leeuwen
G.J.L. Kaspers
C.B. Lambalk
E. van Dulmen-den Broeder

Cancer Treat Rev.
2017 Feb;53:10-24.

ABSTRACT

Background

Anti-cancer treatment may reduce the fertile life span and induce premature menopause. This review aims to provide an overview of the available literature on effects of chemotherapy only on the incidence of ovarian dysfunction and to evaluate the relationship between dose of chemotherapy, age at time of treatment, and time since treatment in female survivors of childhood and young adult cancer.

Methods

A comprehensive search of electronic databases was performed (search date December 2015).

Results

45 studies were included, describing, in total, 5607 female survivors. Median age at menopause was earlier in cancer survivors than in the general population. The prevalence of amenorrhea varied from 0 to 83%. Those exposed to MVPP protocols were at highest risk for amenorrhea (39-79%), as were breast cancer survivors receiving cyclophosphamide-containing regimens, in whom the prevalence of amenorrhea was 40-80%. The most important risk factors for ovarian dysfunction were: (1) alkylating agents, specifically procarbazine and busulfan, (2) older age at treatment.

Conclusion

Breast cancer survivors, those treated with procarbazine or other alkylating agents and those with a higher age at diagnosis are at highest risk of diminished ovarian function. However, all studies included in this review showed methodological limitations. It is imperative that nation-wide registries guarantee long-term follow-up during the adult life of cancer survivors.

BACKGROUND

Over the last decades, more effective treatment regimens for childhood cancer have led to a significant rise in long-term survival. However, survivors treated for childhood and young adult cancer before the age of 40 years, are at risk of adverse late effects on the reproductive system, which may not become apparent until many years later.

All women are endowed with their entire non-replenishable pool of follicles from birth. From the age of approximately 31 years fecundity decreases, leading to natural sterility, menstrual cycle irregularity and menopause [3]. Age at natural menopause in the developed world ranges between 40 years and 58 years. Other factors explaining the variance in age at menopause, include both ethnicity as well as lifestyle and reproductive factors (smoking, parity, age at menarche) [2]. Because the pool of primordial follicles in women is fixed, cancer treatment may substantially deplete this oocyte pool through its associated cell kill. This potentially leads to a reduced ovarian reserve, infertility and premature menopause.

It has been demonstrated that radiotherapy to the lower abdomen and pelvis as well as total body irradiation have deleterious effects on ovarian function [49]. The degree of harm is related to the localisation, total radiation dose, fractionation schedule and age at time of treatment. However, in contrast to these adverse late effects of radiotherapy on ovarian function, the exact mechanism by which specific chemotherapeutics do harm in females remains unclear. Small studies assessing clinical outcomes of gonadal function, such as AMH and FSH levels and AFCs [56 59-61] and large cohort studies [45 53 62-65] in which gonadotoxicity was assessed by self-report have to a greater or lesser extent demonstrated that exposure to the following chemotherapeutic agents is associated with gonadotoxicity: alkylating agents, procarbazine, antimetabolites, platinum agents, taxanes, and, possibly vinca-alkaloids. It is known that chemotherapy may induce damage to the ovaries, and that this damage is drug-, dose- and age-dependent. Histopathology of the ovary has shown that certain types of chemotherapy destroy the primordial follicles and can lead to ovarian atrophy in animal models. Other mechanisms reported to cause damage to the ovary include injury to blood vessels and focal ovarian cortical fibrosis [66].

In the ASCO recommendation on fertility preservation [67] as well as in the recent updated recommendation of Lambertini et al. [68], high dose radiotherapy to the ovaries, as well as treatment cycles of CMF, CEF, CAF and TAC in women over 40 years are stated as the highest risk for ovarian failure. Over time, treatment protocols have changed from high-dose radiotherapy to multimodal treatment (combining radiotherapy and chemotherapy). The newest protocols contain mostly multi-agent chemotherapy and as little radiotherapy as possible. Therefore, there is a need to determine the effect of chemotherapy without radiotherapy on ovarian function. The aim of this review is to summarize the available literature on the association between chemotherapy and ovarian function in childhood or young adult survivors of cancer. The secondary objective is to evaluate the relationship between ovarian function and specific chemotherapeutic agents, dose of chemotherapy, age at time of treatment, and time since treatment.

METHODS

We included all study designs examining the effect of chemotherapy on ovarian function outcomes, except case reports and case series. Only female survivors of childhood or young adult cancer under the age of 40 years at time of outcome assessment were included. This age cut-off was chosen because premature ovarian insufficiency is defined as having an impaired ovarian function before the age of 40 [69]. For the outcome age at menopause, we did not impose an age restriction at the time of outcome assessment. Minimal follow-up time had to be at least two years after diagnosis. All types of cancer diagnosed were included.

Since the current review aimed to assess the effect of chemotherapy only on the ovary, we excluded all cases in which chemotherapy was unequivocally combined with abdominal or pelvic radiotherapy, total body irradiation or radiotherapy on the brain from the analyses. Additionally, women who underwent surgery to the ovaries were excluded.

Ovarian function was operationalized as attained age at menopause or signs of premature ovarian insufficiency (elevated FSH, amenorrhea before age 40). Secondary outcomes were other signs impaired ovarian function, such as low AMH, low inhibin B, and low AFC. W applied cut-off values as defined by the authors of the original study.

We searched PubMed, EMBASE and CENTRAL (search date December 2015) (Appendix A). We handsearched the reference lists and scanned conference proceedings and trial registers. Two reviewers (AO, MvdB) independently selected articles meeting the inclusion criteria based on the abstract and obtained in full any article, which seemed to meet these inclusion criteria for closer inspection. Items were abstracted from the selected articles in a standardized fashion (Appendix B). Risk of bias was assessed by means of validated checklists according to evidence-based medicine criteria [70 71] (Appendix C). If necessary, we contacted authors of individual studies to clarify or gather missing data. Heterogeneity was assessed both by visual inspection of the forest plots and by a formal statistical test for heterogeneity, i.e. the I^2 statistic [72]. To construct plots for the prevalence of amenorrhea and elevated FSH levels we used the generic inverse variance function of RevMan.

RESULTS

The search yielded 3,178 titles; 203 of which were considered as potentially eligible for inclusion in this review after screening the titles and abstracts of the studies and were thus obtained in full text (Figure 1). 120 were excluded after closer inspection (Appendix D). Of the remaining 83 studies, 17 reported original data that could directly be used for the purpose of this review. The remaining 66 studies did not include the specific information that we needed in the text of the article, but were likely to have collected the necessary data. These authors were approached and asked to provide additional results or send the raw data. 18 out of 66 authors were able to provide the information. The remaining 38 did not answer to our call, were not able to retrieve the data, appeared not to have the data or were not willing

to contribute their data. These articles are described in Appendix D. Three studies could not be included because data on our primary outcome was not mentioned in the specific subgroup we defined (age under 40 years, treated with chemotherapy only) [53 63 73] (Appendix E). However, these studies did build a multivariate model in which radiotherapy was included as a separate variable in the model. Therefore, these multivariable risk models did contain chemotherapy and age at treatment as independent variables and thus a calculated risk measurement for these items. Five additional eligible studies were identified from conference proceedings and one from trial registers. As these studies are still ongoing, only descriptions of these studies were added to Appendix F (Characteristics of ongoing studies). In total, our search identified 45 eligible studies [17-19 40 46 47 54-57 60 74-107]. Characteristics of the included studies are summarised below and their baseline characteristics are described in Appendix G (Characteristics of included studies).

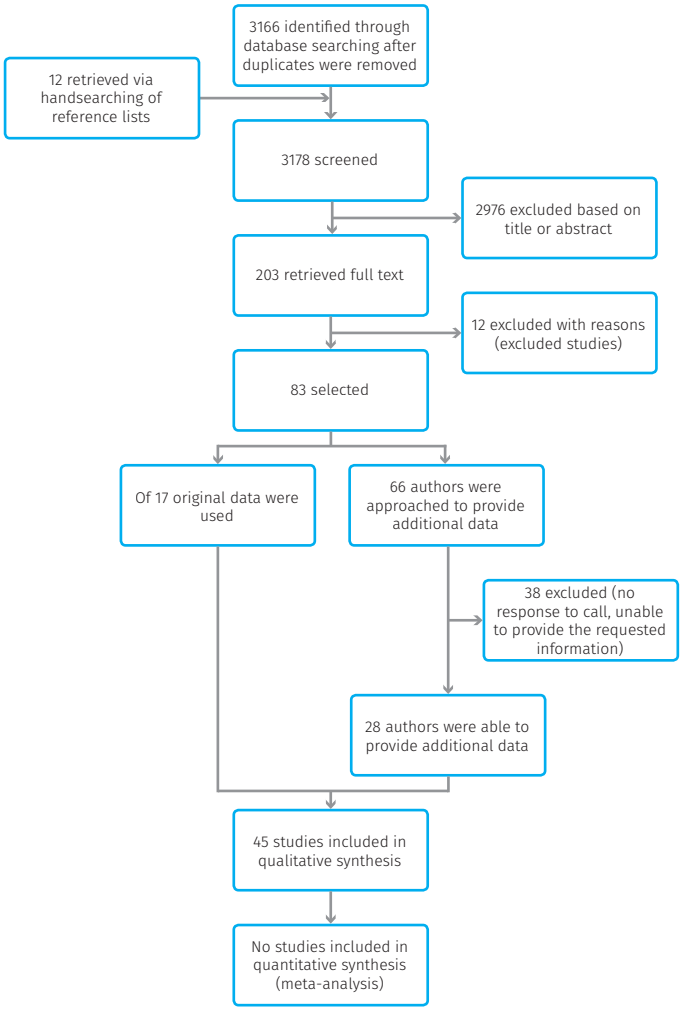


Figure 1 Study flow diagram.

Design

Most studies were designed as retrospective cohort studies [19 40 47 56 74-78 80-82 84-86 88 92-108], cross-sectional studies [17 18 54 55 57 60 91], or longitudinal prospective studies [79 90]. Only Giuseppe et al. conducted a randomized controlled trial [83], whereas Krawczuk et al. conducted a case-control study [89]. Most studies were performed in Europe [17-19 40 46 56 57 60 74-80 82-84 86 89-91 93 95 97 98 101-107], seven were performed in the USA and Canada [46 54 55 92 96 99 100] and six in other countries (Australia (n=1), Brazil (n=1), Israel (n=2) and Japan (n=2))[81 82 85 88 94 108].

Participants

In total, the 45 included studies described 5,607 women. The smallest and largest study population consisted of 6 and 2,127 women, respectively. Follow-up time was reported in 36 studies and ranged from 0.3 years to 48.4 years. Survivors were treated between 1945 and 2012, but in eight studies the time period in which survivors were treated was not mentioned [40 55 57 75 89 91 100 102].

Groups by diagnosis

Twenty-one articles described 3,939 women with HL (Hodgkin lymphoma) or NHL (non-Hodgkin lymphoma) [17 19 40 54 74-76 78-80 82-84 86 89 96 101 103 105 107]. These survivors were aged 8-48 years at diagnosis and 21-60 years at time of study. Thirteen articles reported on 1,302 CCS (childhood cancer survivors) (including various diagnoses) in total [18 46 47 55 57 60 78 90 91 94 95 100 102]. Median age at diagnosis in these studies ranged from 4 years to 17.7 years and age at time of study varied between 12.6 to 40.7 years. Four articles reported on 149 survivors with breast cancer, but did not report on age at diagnosis [85 88 92 93]. Survivors were between 32.5 and 48 years old at time of the study [85 88 92]. Three articles described in total 150 survivors suffering from hematological malignancies (acute lymphatic leukaemia (ALL (acute lymphatic leukemia), AML, HL and NHL) [56 77 104]. Three studies described 31 survivors treated for AML (acute myeloid leukaemia)[81 98 108]. Age at AML diagnosis was between 1 and 30.9 years old. Three studies described 29 survivors treated for osteosarcoma [78 97 106].

Exposure

The individual types of chemotherapy the study participants received as a part of their treatment of childhood or young adult cancer was mentioned in all but four studies [46 54 56 91]. Byrne et al. and Lie Fong et al. specified chemotherapy treatment as containing alkylating agents or not containing alkylating agents and Charpentier et al. reported on the CED (cyclophosphamide equivalent dose; a unit for quantifying alkylating agent exposure independent of study population) [109]. The cumulative dose of the various chemotherapy types was described in twenty-five studies [17-19 40 57 75 77 80 83 85 88-90 92 95-100 102-104 106 108].

In all but two studies (anthracycline-based regimen and methotrexate [85] and "not specified" [91]) the chemotherapy regimens contained alkylating agents. Regimens including procarbazine were reported in all studies that included survivors with HL [17 40 54 57 74-76 78-80 82-84 86 89 90 96 99 101 103 105 107].

Effects of interventions

Age at menopause

Four studies reported on age at menopause [47 77 80 86]. Thomas-Teinturier et al. described 706 CCS of whom 97 had reached menopause (either surgical or non-surgical). Median age at menopause was 44 years (range 18-55). Risk factors for early menopause were exposure to alkylating agents, procarbazine dose and cyclophosphamide dose and treatment after puberty. Risk factors for premature menopause were age at diagnosis, cyclophosphamide dose and exposure to melphalan [47]. De Bruin et al. described 286 HL survivors of whom 65 experienced a premature menopause at a median age of 33.5 years (range 19-39). Higher dose of procarbazine and older age at diagnosis was associated with a shorter time to menopause. King et al. described 18 survivors who were given 6-9 cycles of MVPP. Of these women, 10 reached menopause prematurely. Median age at menopause was not reported, however the range was 30-43 years. In those who were treated with MVPP before the age of 30 years, median time to menopause was 7 years, while in those treated after 30 years of age, median time to menopause was 2 years. Bresters et al. evaluated age at menopause in a group of 109 CCS who received HSCT (hematopoietic stem cell transplantation) with a conditioning regimen containing either busulfan or other alkylating agents. These survivors were treated for different diseases, however, 64% (n=69) suffered from a hematological malignancy. In the whole group, 56% of women entered menopause prematurely, i.e. at a median age of 14.4 years (range 1.0-25.5). Several well-designed studies developed risk models for premature menopause [46 53 101 103], but failed to report age at menopause in the exposed group.

Signs of premature ovarian insufficiency

Prevalence of amenorrhea

The prevalence of amenorrhea after treatment with chemotherapy was reported in twenty-seven studies and 2,975 survivors. Table 1 shows the number of included survivors, mean/median age at diagnosis, mean/median attained age at time of study, the chemotherapy regimen and the reported prevalence of amenorrhea, stratified by diagnosis group. Figure 2 shows that the prevalence varied widely, i.e. from 0 to 83%. Because unexplained heterogeneity was detected (I^2 = 96%), no pooling of results was performed.

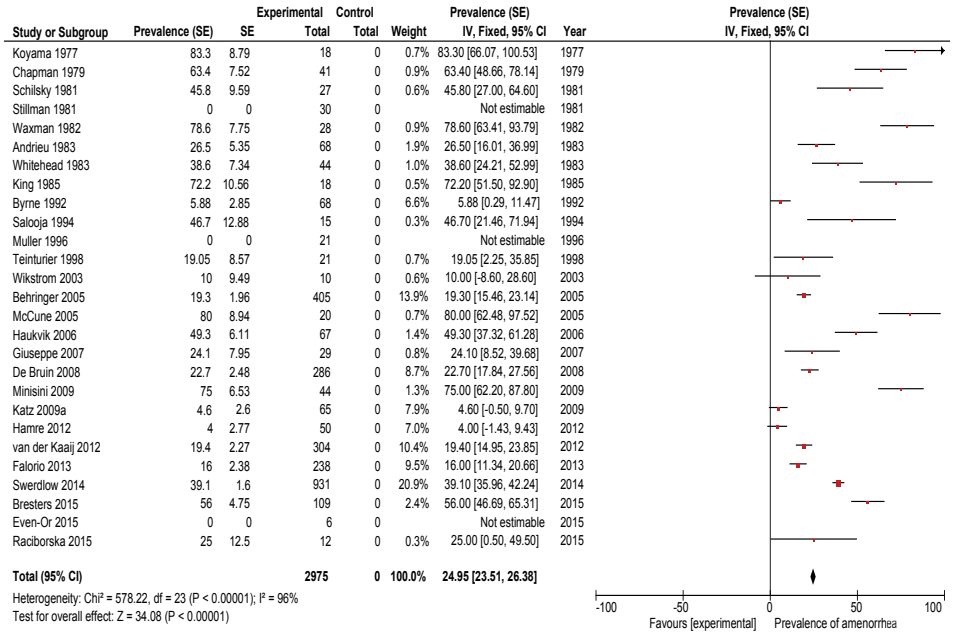


Figure 2 Forest plot depicting the prevalence of amenorrhea prevalence in all the included studies

Prevalence of amenorrhea per diagnosis group

In the twelve studies that described survivors with HL (n=2,245), the prevalence of amenorrhea ranged from 19 to 79% (Figure 3).

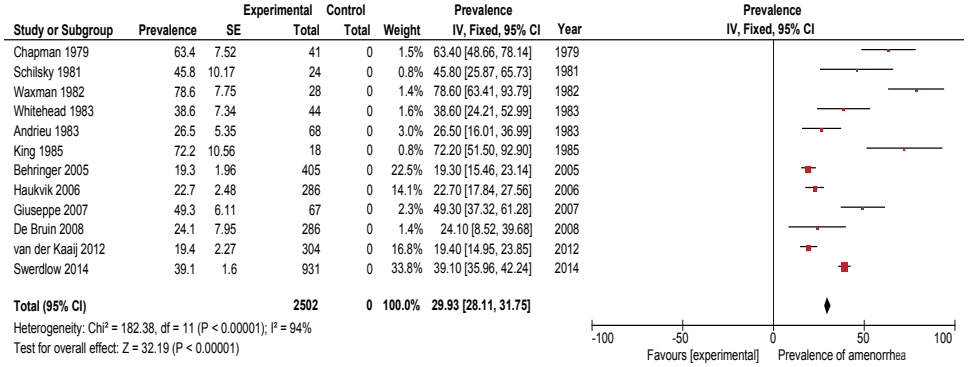


Figure 3 Forest plot showing the prevalence of amenorrhea prevalence in Hodgkin's lymphoma survivors

Also in this subgroup, heterogeneity was detected (I^2=94%), and data could not be pooled. The older studies were more likely to have introduced bias in their studies, as the total cohort from which the study population was derived, was often not described. In addition, in these studies possible confounders were not taken into account [40 74 79 86 99 105]. In six studies MVPP was evaluated. After 3-16 cycles of MVPP, amenorrhea rates were between 39-79% [40 79 86 99 101 105]. In those studies

that assessed the effect of MOPP, amenorrhea rates of 24.5 – 63% were reported [74 80 101 103]. In the study of de Bruin et al., chemotherapeutic agents were divided into separate agents or classes. It seemed that increasing doses of procarbazine could explain high rates of amenorrhea (HR 8.1 (95%CI 2.0-32.8)) whereas mechlorethamine could not [80]. ABVD was found to contribute very little to the risk of amenorrhea (rates reported 0 to 3.8%) [75 80 101].
In the studies reporting on survivors of heterogeneous forms of childhood cancer (n=1,139), the prevalence of amenorrhea was significantly lower compared to Hodgkin lymphoma survivors. The rate varied from 0 to 19%. For breast cancer survivors, a wide range was found (4-80%). However, this wide range was primarily caused by the inclusion of the study by Katz et al. This was the only study in which survivors did not receive cyclophosphamide as part of treatment. In the three studies in which breast cancer survivors did receive cyclophosphamide-containing regimens, the prevalence of amenorrhea ranged from 40 to 80% [88 92 93]. After treatment for AML, the prevalence of amenorrhea after treatment varied widely. Although Salooja et al. and Bresters et al. found comparable rates between 47-56% [77 98], Even-Or et al. found that none of the six studied survivors became amenorrheic after treatment with only melphalan as a conditioning regimen [81]. Two studies reporting on 22 osteosarcoma survivors reported the prevalence of amenorrhea to range between 10-25% [97 106].

Prevalence of elevated FSH levels

The prevalence of elevated FSH levels was reported in eight studies [60 78 79 81 96 102 106 107] and rates varied from 8 to 83% (Figure 4). Again, unexplained heterogeneity was detected (I^2 = 95%), therefore no pooling of the data could be performed. The prevalence rates of elevated FSH levels was 8-74% in three studies on childhood cancer [60 78 102], 20-83% in four studies on HL and NHL [78 79 96 107], 50% in one study on AML [81] and 0 to 40% in two studies on osteosarcoma [78 106].

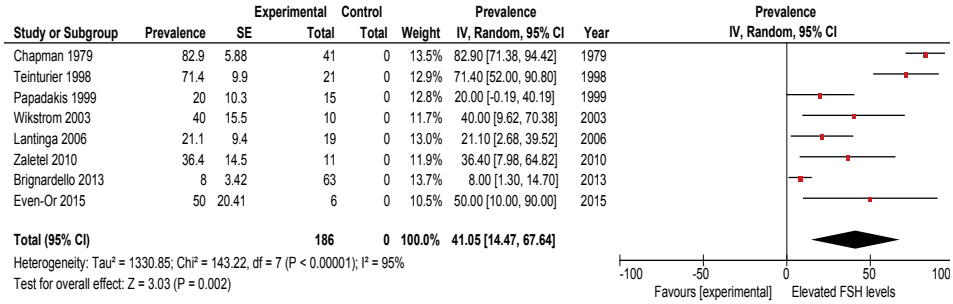


Figure 4 Forest plot depicting prevalence of elevated FSH levels

Table 1 Characteristics of studies reporting on the rate of amenorrhea after treatment with cytotoxic therapy

Reference	n	Treated in childhood/as young adult	Age at diagnosis	Attained age
Lymphoma (HL and NHL)	5128			
Swerdlow 2014[101]	2127	Both	5.5% between 0-14 years 20.7% between 15-19 years 29.2% between 20-24 years 24.3% between 25-29 years 20.2% between 30-35 years	n/m
Van der Kaaij 2012 [103]	1700*	Young adult	28.0 years[$]	43.5 years[$]
Behringer 2005 [75]	405	Young adult	< 30 years n = 245 30-40: n = 160	n/m
De Bruin 2008 [80]	286	Both	14-39 years (14-21 32.2%; 22-28 35.3%; 29-39 32.5%)[$]	25.9 ± 6.6[$]
Falorio 2013 [82]	238	Both	29 years (range 15-39)[$]	74% < 30 years, 26% > 30 years

CT regimen	Definition of amenorrhea	Amenorrhea
		19 – 79%
Type of chemotherapy: ABVD, ChlVPP, LOPP, MVPP, MOPP, BEAM+classic AA, BEAM without AA. Surgery: n/s Radiotherapy other than subdiaphragmatic and brain: n/s (however subgroup analyses of those who received pelvic RT and those who received at least 30 Gy to the brain were excluded)	Not specified Excluded: Sterilizing surgery Unknown menstrual cycle due to contraceptive use	Before 40 years of age: 364/931 (39.1%) with classic AA, no pelvic RT 2/144 (1.4%) ABVD 150/383 (39.2%) ChlVPP 95/273 (34.8%) LOPP 66/115 (57.4%) MVPP 38/91 (41.8%) MOPP 37/53 (69.8%) BEAM + alkylating agent 1/5 (20%) BEAM but no other classical AA
Vinblastine, procarbazine, MOPP (mechlorethamine, vincristine, procarbazine, prednisone), ABV (doxorubicin, bleomycin, vinblastine), BEACOPP (cyclophosphamide, doxorubicin, vincristine, bleomycin, etoposide, procarbazine, prednisone), ABVD (doxorubicin, bleomycin, vinblastine, dacarbazine), EBVP (epirubicin, bleomycin, vinblastine, prednisone)	Cessation of menstrual cycle > 1 year Excluded: Sterilizing surgery Unknown menstrual cycle due to contraceptive use	POF 59/304 (19.4%) 1.5-3 cycles of MOPP 44/155 (28%) > 4 cycles of MOPP 25/47 (53%)
ABVD; COPP/ABVD; BEACOPP standard; dose-escalated BEACOPP	No menstrual bleeding after therapy and at the time of the survey (median observation time was 3.2 years after treatment, min 7 months, max 6.3 years)	Overall 78 (19.3%) Per treatment protocol: HD7B (ABVD) 1/26 (3.8%) HD8 (COPP/ABVD) 13/188 (6.9%) HD9A (COPP/ABVD) 12/32 (37.5) HD9B (BEACOPP) 12/53 (22.6) HD9C (dose esc BEACOPP) 38/74 (51.4)
MOPP (n=94; 32%), MOPP/ABV (n=60; 21.0%), ABVD (n=19; 6.6%), EBVP (n=23; 8.0%)[$] ABVD/VBM (n= 19), BEACOPP / COPPEBVCAD (n=13), escalated BEACOPP (n=5)[$]	Cessation of menses before age 40 years Impaired gonadal function: irreversible amenorrhea after treatment or if they failed to conceive despite pursuing pregnancy with regular menses. Irreversible amenorrhea: absence of resumption of normal menstrual cycles in the 6 months subsequent to the conclusion of treatment	65/286 (22.7%) premature menopause[$] 37/238 (16%)[$] Of those with amenorrhea no subgroups were described, but impaired gonadal function was reported in 21% of survivors with ABVD/VBM, 35% COPPEBVCAD/BEACOPP and 43% with esc BEACOPP

Table 1 - Continued

Reference	n	Treated in childhood/as young adult	Age at diagnosis	Attained age
Lymphoma (HL and NHL)	5128			
Andrieu 1983 [74]	68	Both	Median 28 years, (16-45 years) in group with 3 cycles of MOPP, median 27 (range 17-40 in group with 6 cycles of MOPP	n/m
Haukvik 2006 [84]	67	Both	37% between 9-29 years, 38% between 30-40 years	44 years
Hamre 2012 [110]	50	Childhood	15.1 years (2.2-18.0)*	29.1 years (19.9-39.9)[$]
Whitehead 1983 [40]	44	Both	23 years (13-45)	n/m
Chapman 1979 [79]	41	Young adult	n/m	Median 30 years (20-51)
Giuseppe 2007 [83]	29	Both	24.3 years ± 7.9	30.26 years ± 8.71
Waxman 1982 [105]	28	Both	Median 30 years (17-48)	Median 39 years (25-60)
Schilsky 1981 [99]	27	Both	16-46 years	30 years (18-46)
King 1985 [86]	18	Both	17-39 years	n/m

CT regimen	Definition of amenorrhea	Amenorrhea
		19 – 79%
MOPP either 3 cycles (n=49) or 6 cycles (n=19)	Not specified	2/42 (4.8%) in those < 30 years of age at treatment 16/26 (61.6%) in those > 30 years of age at treatment 3 cycles of MOPP 24.5% 6 cycles of MOPP 31.5%
Mustine, vincristine, procarbazine or ChlVPP (chlorambucil, vinblastine, procarbazine, prednisone); ABVD	Persistent amenorrhea before the age of 41, after other possible causes for this amenorrhea have been excluded (i.e. hysterectomy)	CT-group overall 33/67 (49%) Without alkylating agents: 3/13 (23%) with alkylating agents 30/54 (55.6%)
ABVD, EBVP, CHOP/COP, BFM 90/93, cyclophosphamide, MVPP, ChlVVP, OEPA/ OPPA, COPP, HDT, TBI, BEAM, LVPP*	Menopause: no menstruation in preceding 2 years Amenorrhea: no menstruation during the last 90 days	2/50 (4%) of women < 40 years[$] No specification of treatment groups
MVPP, MOPP, additional chemotherapy (bleomycine, adriamycin, vincristine, CCNU in 1 patient)	Not specified	17/44 (39%)
MVPP of ChlVPP ± doxorubicine, bleomycine, cyclophosphamide, chlorambucil	No menstruation for minimal 6 months and maximum 7 years	26/41 (63.4%)
MOPP/ABVD/DHAP	No menstruation (mean 5.9 years, SD 4.5 years)	7 (46%) No specification of treatment groups
MVPP 6-16 cycles	No menstruation > 3 months after end of treatment	22/28 (78.6%) In those aged < 30 years 12/17 (71%) in those aged > 30 years 10/11 (91%)
MOPP/ MVPP + RT mantle and peri-aortic	Cessation of menses of greater than 6 months duration	11/24 (46%) Age at treatment > 25 years : 89% Age at treatment < 25 years: 20%
MVPP 6-9 cycles	Raised concentrations of FSH and LH	13/18 (72%), of which 10 premature (< 40 years): 56%

Table 1 - Continued

Childhood cancer	1139			
Byrne 1992 [46]	1067	Childhood	Mean 13.6 years	Mean 32.3 years
Stillman 1981 [100]	30	Childhood	6.9 years (0.1-17)	n/m
Muller 1996 [95]	21	Childhood	12.9 years (3.5-16.5)	22.5 years (18.1-28.8)
Teinturier 1998 [102]	21	Childhood	9 years (2-17)	14.5 years (11.5-21)
Breast cancer	248			
Minisini 2009 [93]	145	Young adult	n/m	< 40 years
Katz 2009 [85]	65	Young adult	n/m	32.5 ± 4 years (20.3-38.5)
McCune 2005 [92]	20	Young adult	n/m	36.7 years (29.8-41.0)
Koyama 1977 [88]	18	Young adult	n/m	33-48 years (mean 42.4)

		0-19%
Alkylating agents	Absence of menses Not considered: secondary amenorrhea, temporary cessation of menses, menstrual problems, surgical menopause	CT only: 4/68 (5.9%) RT above diaphragm + CT: 3/38 (7.9%)
Actinomycin D 4.3 mg, chlorambucil 1665 mg, cyclophosphamide 29025 mg, methothrexate 4717 mg, vincristine 45 mg, nitrogen mustard 20mg	Ovarian failure: amenorrhea accompanied by persistent elevation of both serum gonadotropins: FSH > 40 IU/l and LH > 25 IU/l, and FSH/LH ratio > 1	0 / 30
Cyclophosphamide 2.8 (2.0-14.0)g/m^2, ifosfamide 36 (8-72) g/m^2, CCNU 0.47 (0.28-0.66) g/m^2, procarbazine	Not specified	0/21
Cyclophophamide, ifosfamide, procarbazine, lomustine, teniposide, melphalan, carmustine, carboplatin, vincristine, etoposide, aracytine, busulfan	No menstruation at least 3 years	4/21 (19%)
		Treated with cyclophosphamide: 40 - 80% Treated without cyclophosphamide: 4.6%
CMF 35.2% (51/145 pts); anthracyclines 87.6% (127/145); taxanes 42.8 (62/145)[$]	Cessation of menses ≥ 6 months	76.2%[$]
MTX or antracycline-based CT	No regular menses	n=3, 4.6% [$]
Cyclophosphamide, doxorubicin Cyclophosphamide: 9 g/m^2 (2.4-14.45)	No pregnancy during the study interval, > 3 months cessation of menses, decreased estradiol levels (< 20 pg/ml) and elevated FSH levels (≥ 30 IU/L)	16/20 (80%)
Cyclophosphamide daily oral dose of 100 mg for 3-14 months (total dose 8.4-39.9 gram).	Not specified	15/18 (83.3%) < 30 years 3/5 (60%) (cum dose 21-53,2 g CY) 30-40 years 2/5 (40%) (cum dose 10.5-20,2 g CY) > 40 13/13 (100%) (cum dose 8.4-39.9 gram CY)

Table 1 - Continued

Reference	n	Treated in childhood/as young adult	Age at diagnosis	Attained age
AML*	132			
Bresters 2015 [77]	109	Childhood	9.3 years (0.3-18.9)	21.4 years (10.9-45.2)
Salooja 1994 [98]	15	Young adult	30.9 years ± 5.7	n/m
Even-Or 2015 [81]	8	Childhood	10.5 years (1-19)	25.5 years (13-36)
Osteosarcoma	22			
Raciborska 2015 [97]	12	Childhood	13.9 years (2.8-17.7)	16.3 years (10.5-31.6)
Wikstrom 2003 [106]	10	Childhood	12.9 years (6-15)	18.6 years (16-22)

** Data abstracted from the original data*
$ Additional data (from a subgroup of the study population) received from the author after enquiry

CT regimen	Definition of amenorrhea	Amenorrhea
		0-56%
Conditioning regimen before HSCT: with busulfan or without busulfan (but with cyclophosphamide/melphalan/ifosfamide/ treosulfan/thiotepa) Cumulative dose: busulfan 8-200 mg/kg, cyclophosphamide 100-200mg/kg, melphalan 140 mg/m², ifosfamide 2-6 g/m², treosulfan 42 g/m², thiotepa 8 mg/kg	Elevated gonadotropin levels (FSL and LH ≥ 10 IU/l) and low estradiol levels (< 40 pmol/l) and in pre-pubertal females the absence of spontaneous pubertal development after the age of 12 years or in pubertal or post-pubertal females the absence of menses without spontaneous recovery	61/109 (56%) Without busulfan 5/17 (29%) With busulfan 23/34 (68%)
Cyclophosphamide 4.5 g/m², daunorubicine 50mg/m², carmustin 300 mg/m², ARA-C 800 mg/m², 6-thioguanine 800 mg/m²	At least 10 months without menses	7/15 (46.7%)
Melphalan and hematopoietic stem cell transplantation (HSCT)	Not specified	0/6
		10-25%
Vincristine, ifosfamide, doxorubicine, etoposide (VIDE) ± VAI (vincristine, actinomycine D, ifosfamide) / VAC (vincristine, actinomycine D, cyclofosfamide) Cumulative dose: VIDE/VAI: ifosfamide 102 g/m², doxorubicin 360 mg/m², etoposide 2,700 mg/m², VIDE/VAC: ifosfamide 60 g/m², doxorubicin 360 mg/m², etoposide 2,700 mg/m², cyclophosphamide 10.5 g/m²	Premature ovarian failure: at least two of the following: (1) amenorrhea for at least 4 months; (2) serum FSH > 25 U/L on two occasions; (3) low serum estradiol at the same time as FSH was elevated Primary amenorrhea: absence of menses at age 16 years in the presence of normal growth and secondary sexual characteristics, or at age 13 years if no menses had occurred an there was an absence of secondary sexual characteristics	3/12 (25%) with premature ovarian failure
Methotrexate, cyclophosphamide, ifosfamide, etoposide, cisplating, doxorubicin, bleomycin, dactinomycin, vincristine Treated according to specified protocol for malignancy:CCC782, ISG/SSG I, SSG II and a modified Rosen T10	Not specified	1/10 (10%)

Table 2 Characteristics of studies reporting on mean or median FSH levels after treatment with cytotoxic therapy

Reference	N	Treated in childhood/ as young adult	Diagnosis	Age at diagnosis	Age at time of study
Behringer 2012	106	Young adult	HL	28 ± 7 years	32 years ± 7
Giuseppe 2007	29	Young adult	HL	24.3 years ± 7.9	30.3 years ± 8.7
Waxman 1982	28	Young adult	HL	Median 30 years (17-48)	Median 39 years (25-60)
Krawczuk 2014	12	Both	HL	15.2 years ±2.6	21.4 years ± 4.4
Brignardello 2013	63	Childhood	ALL, HL, NHL, AML, brain tumors, sarcomas, other	10.0 years ± 7.3	24.89 years (7-40)
Thomas-Teinturier 2015	105	Childhood	ALL, HL, NHL, sarcoma, neuroblastoma, other	9.3 years (0.04-17.7)	25 years (17-40.7)
Gracia 2012	65	Both	HL, NHL, leukemia, sarcoma, Wilms, breast, other	11 years (4 months-33)	25.6 years (15-39)
Martin 2009	43	Childhood	ALL, NHL, Wilms, bone, germ cell, LCH, neuroblastomas, PNET, liver tumor	4.1 years ± 2.9	12.7 years ± 3.0
Krawczuk 2013	33	Childhood	HL, Wilms tumor, CML, AML, ALL, soft tissue sarcoma, NHL, neuroblastoma	14 years ± 3.5 7.6 years ± 4.7	23.6 years ± 4.0 18.2 years ± 2.7

Chemotherapy	Cycle day on which blood sample was taken	FSH level (IU/l)
ABVD	End of pill break or at day 3 of menstrual	3.0
BEACOPP+ABVD	cycle	4.3
MOPP/ABVD/DHAP	Day 3 of the menstrual cycle Amenorrheic patients: at first visit or after three months suspension of hormonal replacement therapy	11.2 ± 20.3
MVPP 6-16 cycles	Day 3-5 of menstrual cycle	Median 18 (1.5-> 25)
3x MVPP + 3x B-DOPA dacarbazine 900 mg/m², procarbazine 3000 mg/m²; nitrogen mustard 36 mg/m²	Day 2-4 of menstrual cycle	7.1 (2.8), Median 7.0
Various	Not specified	Median 4.6 (17-36.6) ALL (n=20): 4.3 (1.7-8.5) HL (n=19) 4.6 (2.3-36.6) Osteosarcoma (n=7) 4.3 (1.7-8.5)
Cumulative dose: cyclophosphamide: median 4.6 g/m² (1-22)(n=71). Ifosfamide median 48 g/m² (3-104)(n=33) Procarbazine median 3 g/m² (0.3-9.1)(n=23)	Regular cycle: day 2-5 of menstrual cycle Asked to refrain from oral contraceptive for at least 1 months On oral contraceptive pills: day 7 of the pill free interval	Median 6.2 (2.1-52.6)
Low AA dose score High AA dose score	Regular cycle: day 1–4 of menstrual cycle Asked to refrain from exogenous hormones for at least 4 weeks: during the subsequent menstrual cycle Irregular cycles or no menses for 6 weeks after stopping hormones: any time.	7.9 (6.6-9.5) 10.6 (8.7-13.0)
Not specified	Follicular phase	Median 5.0 (0.2-20.1)
High risk Median and low risk	Day 2-4 of menstrual cycle	29.4 ± 42.3 7.5 ± 3.0

Table 2 - Continued

Reference	N	Treated in childhood/ as young adult	Diagnosis	Age at diagnosis	Age at time of study
Miyoshi 2008	28	Childhood	ALL, AML, NHL, CML, HL, neuroblastoma, RMS, hepatoblastoma, PNET, Wilms	6.0 years ± 4.5	18.6 years ± 6.5
Lie Fong 2008	13	Both	Hematological cancer	29.4 years (16.0-36.2)[$]	33.1 years (25.7-40.6)[$]
Vatanen 2013	28	Childhood	ALL, AML	9 ± 4.3 years (1-19) (total group n = 92)	22 ± 6.3 years (9-41) (total group n = 92)
Even-Or 2015	6	Childhood	AML	10.5 years (1-19)	25.5 years (13-36)

[$] *Additional data (from a subgroup of the study population) received from the author after enquiry.*

FSH levels

Thirteen studies reported the mean or median FSH levels of 561 survivors in total [55-57 76 78 81 83 89-91 94 104 105]. Ten studies reported that blood samples were taken in the follicular phase [55-57 76 81 83 89-91 105], in three studies, the timing of the sampling was not specified [78 94 104]. Only four studies reported when patients on oral contraceptives were sampled [55 57 76 83]. Due to the heterogeneity of the studies that reported on FSH levels, data could not be pooled and the descriptives of the studies are summarized in Table 2.

Other markers of ovarian function

AMH

Twelve studies reported on AMH levels as a marker for ovarian function in 585 survivors in total [17-19 54-57 75 78 81 83 89 90]. In clinical practice, AMH levels below 1 ng/ml are thought to be indicative of a diminished ovarian reserve, before signs of cycle irregularity, amenorrhea or elevated FSH levels. From six studies, it was possible to assess whether low AMH levels were indeed present in the study population [17-19 55 78 81]. The characteristics, mean/median AMH levels and prevalence rates of low AMH levels are summarized in Table 3.

Inhibin B

Three studies reported on inhibin B levels as a marker of ovarian function after treatment for childhood and young adult cancer [56 83 91]. In HL survivors described in the study of Giuseppe et al., inhibin B levels were 58.2 ng/L ± 78.7 [83]. In survivors of Hematological cancer, Lie Fong et al. reported a median inhibin B level of 25 ng/L (range 9-127) in survivors receiving chemotherapy that included alkylating agents, and a median level of 90 ng/L (range 85-95) in survivors who did not receive

Chemotherapy	Cycle day on which blood sample was taken	FSH level (IU/l)
Various	Not specified	11.4 ± 11.3[$]
Alkylating (n=10) vs Non-alkylating (n=3)[$]	Day 2-5 of the menstrual cycle	Alkylating (n=10) 8.3 (3.8-154)[$] Non-alkylating (n=3) 8.7 (5.9-20.0)[$]
CY only, BU only, BU+CY	Not specified	Median 17.7 mean 37.7 ± 39.9
Melphalan and hematopoietic stem cell transplantation (HSCT)	Day 2-5 of the menstrual cycle	Median 8.9 (2.02 – 13.6)

alkylating agents [56]. Healthy controls in this study had higher levels of inhibin B (113 ng/L (range 4-206)). Martin et al. reported a median inhibin B level of 33 ng/L (range 10-179) in a young study population, diagnosed with various types of cancer [91].

AFC

Four studies reported on AFC as a marker of ovarian function after treatment for childhood and young adult cancer in 152 women in total [56 57 81 83]. In 105 survivors diagnosed with various types of childhood cancer and exposed to alkylating agents (either cyclophosphamide, ifosfamide or procarbazine), median AFCs were 12 (range 1-40), in comparison to 11 (range 8-24) in healthy controls. Moreover, survivors were divided into three groups based on the treatment they received: alkylating agents alone, alkylating agents and subdiaphragmatic radiotherapy or high dose alkylating agents (> 10 g/m^2 cyclophosphamide, > 40 g/m^2 ifosfamide, no cut-off stated for procarbazine). The total number of antral follicles was significantly reduced only in the survivors exposed to high dose alkylating agents. HL survivors had significantly lower AFC than survivors of ALL [57]. Six survivors with AML receiving melphalan as a conditioning regimen before HSCT had AFC varying from 0 to 4, with all of these survivors reporting to have regular cycles [81]. Giuseppe et al. compared AFCs of 29 HL survivors who had regular cycles to those who were amenorrheic. The amenorrheic women were significantly older (35.7 vs. 26.3 years) and had lower AFCs (2.3 vs. 4.7). However, as antral follicle counts are known to be strongly associated with age, it is uncertain if the difference between these groups was not solely caused by the difference in age [83]. This was not assessed in a statistical analysis. The mean antral follicle count was 4.0 in the survivors receiving an alkylating agent containing regimen in the study of Lie Fong et al., whereas healthy controls showed a mean antral follicle count of 14 (range 2-24) [56].

Table 3 Characteristics of studies reporting AMH levels after treatment with cytotoxic therapy

Reference	N	Diagnosis	Age at diagnosis	Age at time of study
Behringer 2012	106	HL	28 years ± 7	32 years ± 7
Giuseppe 2007	29	HL	24.3 years ± 7.9	30.3 years ± 8.7
Krawczuk 2014	12	HL	15.2 years ± 2.6	21.4 years ± 4.4
Hamre 2012	50	HL, NHL	15.1 years (2.2-18.0)	29.1 years (19.9-39.9)
van Waas 2011	20	NHL	8 years (2-16)	21 years (9-40)
Charpentier 2014	66	Leukemia, HL, sarcoma	11.5 years (1.8-17.3)	23.3 years (18.2-34.2)
Thomas-Teinturier	105	ALL, HL, NHL, sarcoma, neuroblastoma, other	9.3 years (0.04-17.7)	25 years (17-40.7)
Gracia 2012	65	HL, NHL, leukemia, sarcoma, Wilms, breast, other	11 years (4 months-33 years)	25.6 years (15-39)
Rendtorff 2010	56*	ALL, HL, NHL, osteosarcoma, renal tumors, liver tumors, brain tumors, AML, Ewing sarcoma, germcell tumor, soft tissue sarcoma, neuroblastoma	13 years	25 years (21-38)
Brignardello 2013	37	ALL, HL, NHL, AML, brain tumors, sarcomas, other	10.0 years ± 7.3	24.9 years (7-40)
Krawczyk 2013	33	HL, Wilms tumor, CML, AML, ALL, soft tissue sarcoma, NHL, neuroblastoma	14 years ± 3.45 7.6 years ± 4.7	23.6 years ± 4.0 18.2 ± 2.7
Even-Or 2015	6	AML	10.5 years (1-19)	25.5 years (13-36)

** Data abstracted from the original data*

Chemotherapy	AMH (ng/ml)	Proportion low AMH > 1 ng/ml
ABVD	2.2	n/s
BEACOPP+ABVD	0.9	
MOPP/ABVD/DHAP	1.2 ± 1.6	n/s
3x MVPP + 3x B-DOPA	3.2 ± 2.1 Median 3.5	n/s
Low, medium, high toxic	Low: 3.0 (0.3-4.7) Medium: 1.5 (< 0.2-7.2) High:> 0.2 (< 0.2-0.5)	Low 2/10 (20%) Medium: 8/25 (32%) High: 10/10 (100%)
Various	5.3 (3.0-7.5)	3/22 (13.6%)
CED score	3.6 (0.1-15.1) Leukemia (n=3): 7.9 (0.3-15.1) HL (n=16): 1.8 (0.1-10.3) Sarcoma (n=4): 1.8 (1.4-12.0)	n/s
Cumulative dose: Cyclophosphamide: median 4.6 g/m^2 (1-22)(n=71) Ifosfamide median 48 g/m^2 (3-104)(n=33) Procarbazine median 3 g/m^2 (0.3-9.1)(n=23)	1.5 vs. 3.1 in controls	n/s
Low AA dose score	3.1 (2.2-4.4)	Low AAD score: 0
High AA dose score	0.5 (0.3-0.9)	High AAD score: 100%
Various	n/s	17/56 (30.4%) Very low (AMH < 0.1) in 7/56 (12.5%)
Various	Total group: median 2.7 (0-6.5), mean 2.6 ± 0.8 ALL (n=20, missing n=10) median 2.8 (1.5-3.8), mean 2.6 (0.8) Hodgkin lymphoma (n=11, missing n=8): Median 1.2 (0-5.1), mean 1.8 (1.8) Osteosarcoma (n=5, missing n=2): Median 3.7 (2.1-6.5), mean 4.2 (2.1)	Total group: 8/37 (22%) ALL: 0/20 (0%) HL: 5/11 (45%) Osteosarcoma: 0/5 (0%)
Overall group	1.8 ± 0.9	n/s
High risk	1.0 ± 0.8	
Median and low risk	2.7 ±2.5	
Melphalan and hematopoietic stem cell transplantation (HSCT)	0.2 (0.1-3.0)	5/6 (83.3%)

Risk factors for reduced ovarian function

Twenty-seven studies assessed possible risk factors for the development of fertility-related late effects [18 40 46 54 57 60 74-80 82 84 86 88 91 93 97 99-105], twenty of which used adequate measures of risk estimation [18 46 54 57 60 75-78 80 82 84 91 93 97 100-104].

The risk of bias was deemed to be small in 3 out of 45 studies [80 101 103]. These were well-conducted large cohort studies in which hazard ratios were calculated by means of proportional hazards models. This form of analysis provides an estimate of the effect of the exposure on the risk of menopause, but allows the adjustment for other explanatory variables. Most importantly, it takes into account time as a variable. Hazard ratios and 95% confidence intervals, as stated in these studies, are summarized in Table 4.

Treatment with BEACOPP [75 82], alkylating agents in general [46 47 57 80 82 84 101 103 104], CED-score [54], procarbazine [47 54 57 80], ChlVPP [84], busulfan [77 102 104] (although not significant in the study of Bresters et al.), melphalan [47], etoposide [101], MOPP [103] and CMF-based chemotherapy [93] were found to be risk factors for reduced ovarian function in multivariate analysis. All treatment regimens that were found to have a detrimental effect on ovarian function contained one or more alkylating agents (such as cyclophosphamide, chlorambucil, busulfan or procarbazine). However, in the study of Swerdlow et al., which contained the largest included study population of HL survivors, etoposide was also associated with a higher risk of premature menopause [101]. Minisini et al. showed that previous childbearing was independently associated with a higher risk of reduced ovarian function. Older age at treatment was found to be a risk factor in ten studies [46 47 54 60 75 78 80 82 101 103]. However, three other studies did not find a significant effect of this factor on ovarian function [84 91 97]. Bresters et al. did not assess age at diagnosis but rather pubertal stage at diagnosis and found that more advanced pubertal state was independently associated with reduced ovarian function [77]. This was in agreement with findings of Thomas-Teinturier et al. [47].

Seven older studies did report on risk factors but did not apply adequate measures of risk estimation. These results are thus more likely to have been influenced by confounders and should be interpreted with caution. Six studies found that older age at time of treatment was associated with a higher prevalence of amenorrhea [40 74 86 88 99 105]. Chapman et al. found that older women became amenorrheic after fewer cycles of MVPP than younger women treated with similar chemotherapy [79].

Three studies were not included because they did not meet the inclusion criteria for the primary outcomes. Two of these studies were reports of the Childhood Cancer Survivor Study, a study in one of the world's largest childhood cancer survivor cohorts. Although the studies could not be included for the primary outcomes, as treatment was not disentangled in radiotherapy and chemotherapy separately, the authors did perform a well-designed multivariate regression analysis in which radiotherapy and chemotherapy were included in the multivariate model as separate variables [53 63 73] (Appendix E). Because these variables were included in the model, the results of these analyses described the independent effects of the risk factors that

Table 4 Hazard ratios reported in low risk of bias proportional hazard models

	Swerdlow 2014	De Bruin 2008	Van der Kaaij 2012
No alkylating CT, no pelvic RT	1.0	1.0	1.0
Classic alkylating CT, no pelvic RT	20.2 (12.4-32.9)	6.1 (2.7-13.7)	12.3 (5.9-25.7)
No of cycles	(cycles of classical alkylating CT)	(cycles of MOPP)	(cycles of MOPP)
1-3	13.0 (7.0-24.1)	3.8 (1.8-7.8)	9.8 (4.6-21.0)
4-5	16.0 (9.2-27.9)	4.4 (.5-8.0)	21.1 (9.5-46.9)
6	24.1 (15.0-38.9)		
7-12	21.3 (12.9-35.2)	11.6 (6.5-20.7)	
>12	187.8 (77.5-454.8)		
Chemotherapy regimen			
ABVD	1.3 (0.3-6.2)	0.3 (0.0-1.9)	n/s
BEAM	60.6 (34.1-107.8)	n/s	n/s
ChlVPP	22.7 (13.8-37.4)	n/s	n/s
LOPP	17.9 (10.8-29.9)	n/s	n/s
MVPP	37.1 (21.2-65.0)	n/s	n/s
MOPP	19.3 (10.5-35.3)	5.7 (3.6-9.1)	n/s
Alkylating agents			
Procarbazine y/n	n/s	8.1 (2.0-32.8)	n/s
≤ 4.2 g/m^2	n/s	1.3 (0.2-6.8)	n/s
4.2 – 8.4 g/m^2	n/s	5.2 (1.6-17.1)	n/s
> 8.4 g/m^2	n/s	20.4 (4.4-93.6)	n/s
Cyclophosphamide y/n	n/s	3.5 (2.0-5.9)	n/s
Age at first treatment (years)			1.2 (1.2-1.3)
0-19	1.0	1.0	n/s
20-24	1.6 (1.0-2.6)	2.6 (1.5-4.5)	
25-29	2.0 (1.3-3.1)		
30-35	2.0 (1.3-3.1)	5.2 (2.8-9.4)	

we were interested in for this review. These analyses showed that the alkylating agent score, as well as cyclophosphamide dose of more than 5 grams, attained age and a diagnosis of HL were individual risk factors for a reduced ovarian function.

Risk of bias in included studies

Data on the risk of bias of the 45 included studies are described in Appendix G and are shown graphically in Figure 5.
The internal validity of a study represents the bias that is present in a study and thus indicates how valid the results of a study are. Studies were evaluated for internal validity by assessing selection bias, attrition bias, detection bias and the risk of confounding.
The external validation of a study indicates how well the results of the study could be extrapolated to individual survivors in the population at large. All studies were evaluated for external validity by assessing possible reporting bias. Reporting bias can be ruled out when the study group, the length of follow-up and the outcomes are well defined. There was a wide range in follow-up time, and as the risk of adverse effects on ovarian function increases with age, it could be expected that a higher prevalence of ovarian dysfunction was found in those studies that had a longer follow-up. However, in some studies the follow-up time also varied significantly between subjects, making it difficult to interpret the results correctly.

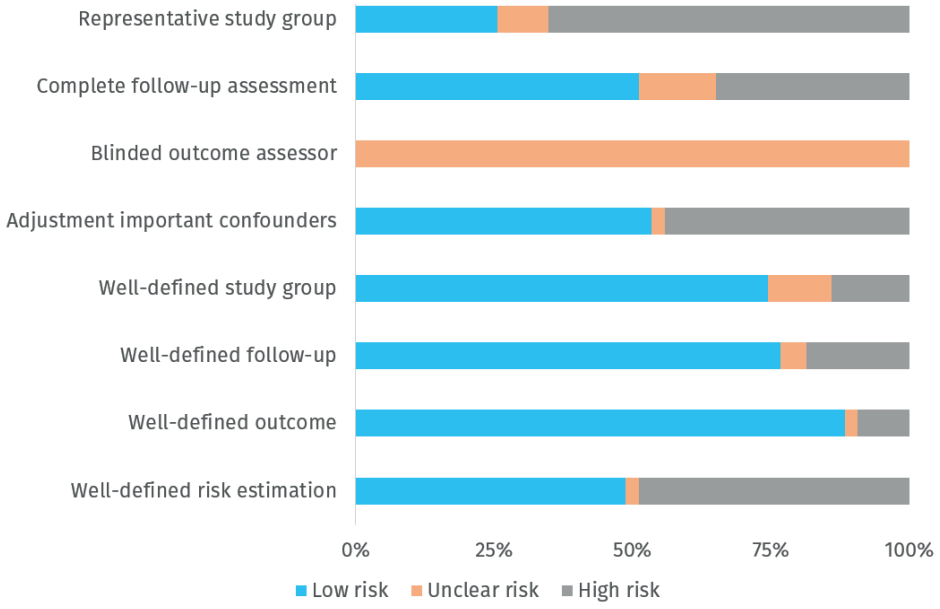


Figure 5 Risk of bias graph, each item presented as percentages across all included studies.

DISCUSSION

For this review, 45 studies met the inclusion criteria including 5,607 survivors in total. Length of follow-up time ranged from 0.3 years to 48.4 years and survivors were diagnosed between 1945 and 2012. Studies were divided into six groups according to diagnosis: (1) HL and NHL (n=21, 3,939 survivors); (2) childhood cancer in general (n=12, 596 survivors); (3) breast cancer (n=4, 149 survivors); (4) hematological malignancies (a combination of NHL, HL, AML and ALL) (n=3, 150 survivors); (5) AML (n=3, 31 survivors); and (6) osteosarcoma (n=3, 29 survivors).

Median age at menopause was 33.5 years in HL survivors. Premature menopause in HL survivors was more frequently present when treatment included procarbazine (in a dose-dependent manner) and prevalence increased with increasing numbers of MVPP cycles. Also, premature menopause was more frequent in survivors with hematologic malignancies exposed to high doses of busulfan as a conditioning regimen for HSCT, with a median age at menopause being as low as 14 years. In the total group of survivors of childhood cancer, age at menopause was earlier than in the general population, but was not necessarily premature (median age 44 years). Of those who had experienced premature menopause, age at diagnosis, cyclophosphamide dose and exposure to melphalan seemed to be the most important risk factors.

The prevalence of amenorrhea after childhood and young adult cancer was reported in twenty-seven studies (2,975 survivors in total) and varied considerably, i.e. from 0 to 83%. It seemed that survivors treated for childhood cancer in general had the lowest risk of amenorrhea (0-20%). However, because of the variety of primary tumours, there was a large heterogeneity in the treatments given. The prevalence of amenorrhea in HL survivors seemed to be highest in those exposed to MVPP (39-79%). However, in HL survivors exposed to ABVD protocols the prevalence of amenorrhea was not higher than the control population (none-3.8%). After MOPP cycles, amenorrhea rates of 25 to 63% were reported. In breast cancer survivors receiving cyclophosphamide-containing regimens, the prevalence of amenorrhea was 40 to 80%. In osteosarcoma survivors, amenorrhea rates varied from 10 to 25%, however, treatment protocols were not comparable. The definition of amenorrhea varied between the studies, making interpretation difficult. Thirteen studies did not specify the definition of amenorrhea, whereas three studies defined it as a cessation of menses > 3 months, one study as > 4 months, five studies as > 6 months, and the remainder longer than at least 10 months. Recurrence of menses is known to occur in some women after treatment; therefore, those studies in which a short absence of menstrual period was acknowledged as amenorrhea might overestimate the actual incidence of permanent amenorrhea.

The prevalence of elevated FSH levels was reported in eight studies and varied from 8 to 83%. Again, unexplained heterogeneity was detected, therefore no pooling of the data could be performed. Because so few studies evaluated elevated FSH levels as an outcome, it was not possible to distinguish the effect of age and type or dose of chemotherapy for this outcome. Not all studies reported the timing of the sampling, which in the case of FSH, is very important, as it varies during the menstrual cycle due to the negative feedback of estradiol. In addition, only four studies reported

when sampling was performed when women were using oral contraceptives. Earlier studies have shown an effect of contraceptive pill use on FSH, when measured on the last day of the pill free week, and even when contraceptive pill use has been stopped in the months prior to blood analysis [111].
In general, FSH levels seemed to be higher in cancer survivors compared to healthy controls, while AMH and inhibin B levels as well as AFC seemed to be lower. This reflects a reduced ovarian function in survivors compared to healthy controls in general. However, only three studies compared the outcomes with healthy, age-matched controls. Because age at time of study varied significantly between the included studies and age is known to be highly associated with ovarian reserve markers, such as FSH, inhibin B, AMH and AFC, it was not possible to pool data and get a more precise insight into the sole effect of chemotherapy on ovarian function. In addition, in most studies the use of oral contraceptives was not corrected for. Earlier studies have shown that contraceptive use may alter endocrinological and ultrasound measures of ovarian function [111].
The most important risk factors for a reduced ovarian function that were identified in more than one study were: (1) alkylating agents, specifically procarbazine and busulfan; and, (2) older age at treatment. In addition, etoposide was found to be an independent risk factor in a large well-conducted study by Swerdlow et al. Although age at treatment is often regarded as a risk factor, one should realize that it is only possible to correctly assess age at menopause when all participants in the study have indeed reached menopause. It might seem that an older age at treatment reduces time to menopause, but it is likely that menopause occurs at the same age as those treated at a younger age, but those who were treated at a younger age, have more time left before menopause occurs (and to be included in a study during this time period).
Due to the explicit inclusion criteria of our review, some well-performed large studies, such as those performed by the CCSS study group could not be included [53 63 73 112]. The fact that cancer treatment is often multimodal, combining chemotherapy with radiotherapy, is responsible for this. Few studies are sufficiently powered to assess the effects of these treatments separately. As we aimed to evaluate the effect of chemotherapy only, we could not include those studies in which radiotherapy was given to the abdomen, brain or to the total body or in which no multivariate model was applied which assessed chemotherapy and radiotherapy as individual factors.
In addition, there was a great diversity in diagnoses, as we included all types of childhood and young adult cancer survivors. Not only did the treatment regimens vary considerably, also patient characteristics such as age at diagnosis varied greatly. Furthermore, over the last decades cancer treatment protocols have changed significantly. At present, radiotherapy can be administered to a smaller volume around the tumour, thereby keeping the radiation effect on the surrounding tissue as low as possible. Moreover, chemotherapy doses have been adjusted to prevent toxicity as much as possible. Therefore, older studies including survivors diagnosed in the forties and fifties of the previous century may have different outcomes than those treated at present. Rates of amenorrhea will thus be much higher in those exposed to cancer treatment half a century ago, and study results should not be extrapolated to the current treatment protocols. As a result, our

primary objective has only been reached in part, i.e. reporting on the prevalence of amenorrhea in survivors with childhood and young adult cancer. Furthermore, it was not possible to pool data in order to assess the effect of different types or doses of chemotherapeutic agents or age at treatment on ovarian function due to high heterogeneity between the included studies.

This review shows that survivors treated for childhood and young adult cancer are at risk for reduced ovarian function, especially after high doses of alkylating agents (and possibly etoposide) and with older age at diagnosis. Careful history taking, in combination with the evaluation of elevated FSH and low AMH levels could be helpful to screen early for ovarian dysfunction. It is however unclear whether markers of ovarian function also predict chance of pregnancy. Physicians should actively advise survivors who are at risk of premature ovarian insufficiency not to postpone motherhood, and should refer survivors to fertility specialists for information on fertility preservation options. Adult females as well as postpubertal girls can be advised to cryopreserve oocytes. For those women who have a partner embryo cryopreservation is an established and successful fertility preservation method [113]. To date, the evidence regarding the effectiveness of ovarian suppression (GnRH-analogs) as a fertility preservation method is still matter of debate, however, meta-analyses show a potential benefit of the therapy in reducing the risk of premature ovarian failure, especially in breast cancer survivors [68]. Ovarian tissue cryopreservation is still experimental. For children, all methods are experimental. Oocyte cryopreservation has been reported in girls aged 1 year and older. Cryopreservation of ovarian cortical tissue has been reported, but no live births have been reported, primarily because of the young age of the study participants [113].

In addition, physicians should warn survivors that although menopause seems imminent, pregnancy is still a possibility and contraceptives should be advised when a pregnancy is not desired. Based on the conclusions drawn from this review, it seems that breast cancer survivors, and those treated with procarbazine and/or high doses of alkylating agents were reported to have the highest risk of premature menopause. To the question whether these high risks are solely due to the treatment given or due to a less favourable starting point (as breast cancer patients tend to be older and thus a lower remaining ovarian reserve), this review has no answer.

APPENDIX

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.ctrv.2016.11.006

3

A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges

A. Overbeek
M.H. van den Berg
L.C. Kremer
M.M. van den Heuvel-Eibrink
W.J.E. Tissing
J.J. Loonen
A.B. Versluys
D. Bresters
G.J.L. Kaspers
C.B. Lambalk
F.E. van Leeuwen
E. van Dulmen-den Broeder
on behalf of the DCOG
LATER-VEVO study group

BMC Cancer.
2012 Aug 23;12(1):363.

ABSTRACT

Background
Advances in childhood cancer treatment over the past decades have significantly improved survival, resulting in a rapidly growing group of survivors. However, both chemo- and radiotherapy may adversely affect reproductive function. This paper describes the design and encountered methodological challenges of a nationwide study in the Netherlands investigating the effects of treatment on reproductive function, ovarian reserve, premature menopause and pregnancy outcomes in female childhood cancer survivors (CCS), the DCOG LATER-VEVO study.

Methods
The study is a retrospective cohort study consisting of two parts: a questionnaire assessing medical, menstrual, and obstetric history, and a clinical assessment evaluating ovarian and uterine function by hormonal analyses and transvaginal ultrasound measurements. The eligible study population consists of adult female 5-year survivors of childhood cancer treated in the Netherlands, whereas the control group consists of age-matched sisters of the participating CCS.

Results
To date, study invitations have been sent to 1611 CCS and 429 sister controls, of which 1215 (75%) and 333 (78%) have responded so far. Of these responders, the majority consented to participate in both parts of the study (53% vs. 65% for CCS and sister controls respectively). Several challenges were encountered involving the study population: dealing with bias due to the differences in characteristics of several types of (non-) participants and finding an adequately sized and well-matched control group. Moreover, the challenges related to the data collection process included: differences in response rates between web-based and paper-based questionnaires, validity of self-reported outcomes, interpretation of clinical measurements of women using hormonal contraceptives, and inter- and intra-observer variation of the ultrasound measurements.

Discussion
The DCOG LATER-VEVO study will provide valuable information about the reproductive potential of pediatric cancer patients as well as long-term survivors of childhood cancer. Other investigators planning to conduct large cohort studies on late effects may encounter similar challenges as those encountered during this study. The solutions to these challenges described in this paper may be useful to these investigators.

INTRODUCTION

Cancer treatment can have detrimental effects on reproductive function. In women, there is evidence that both chemo- and radiotherapy can adversely affect ovarian function, ovarian reserve and uterine function, clinically leading to sub- or infertility, premature menopause and/or adverse pregnancy outcomes [64 114-121]. However, previous studies addressing the late effects of cancer treatment on female fertility have several limitations. Clinical ovarian reserve tests are often lacking (i.e. data from questionnaires only) [63-65 114 116-119 122-128], and study populations are often small and heterogeneous [49 56 59-61 115 129 130]. Therefore, we designed the DCOG LATER-VEVO study (Dutch Childhood Oncology Group - Long term Effects after Childhood Cancer/ Fertility, Ovarian reserve and Premature Menopause (Dutch acronym)) in the Netherlands. Patient inclusion started in 2008.
The study aims to evaluate the effects of cancer treatment on the reproductive system of female childhood cancer survivors (CCS) in the Netherlands and their risk of premature menopause. The effects of treatment in general will be assessed, as well as the effects of different treatment modalities, doses of drugs, radiation sites and doses, and age at time of treatment. The study includes a questionnaire survey and a full panel of ovarian function and reserve tests. The DCOG LATER-VEVO study is the first nationwide childhood cancer survivor study in the Netherlands and during the study period several methodological challenges were encountered.
In this paper the key methodological and practical challenges are discussed as well as the way they were addressed. Other investigators planning to conduct large nationwide cohort studies among childhood cancer survivors will benefit from this information when faced with similar challenges.

METHODS

Design and study population

The DCOG LATER-VEVO study is a multi-center retrospective cohort study including female 5-year survivors of childhood cancer. The study consists of three parts: a questionnaire survey, the provision of a blood sample, and a transvaginal ultrasound measurement of the reproductive organs, the latter two requiring a visit to the outpatient clinic. Approval was obtained from the relevant medical ethics committees and written informed consent was obtained from all participants.
Eligible cohort members are selected from a cohort of patients treated for childhood cancer between 1963 and 2002 at one of the seven Dutch pediatric oncology - and stem cell transplant centers, collectively known as the Dutch Childhood Oncology Group - Longterm Effects after Childhood Cancer (DCOG LATER). This group has developed a nationwide electronic database including patient and treatment details of all CCS in the Netherlands (DCOG LATER database). The study population consists of those female CCS who were treated for a malignancy or central nervous system tumour before the age of 18, who survived for at least 5 years after diagnosis, and who were at least 18 years at study entry (n = 2,331). The exclusion criteria for participation in the study include: deceased before the start of the study (n = 271),

living abroad or unknown address (n = 75), not being able to speak or read Dutch (n = 1), having severe mental sequelae (n = 40), being treated for second malignant neoplasm at the time of study inclusion (n = 34), and previously having indicated not willing to participate in research (n = 13). Thus, a total of 1,897 female childhood cancer survivors are eligible for participation in the DCOG LATER-VEVO study.
Sisters of participating CCS who have never been diagnosed with cancer, who are able to read and speak Dutch, and who are 18 years or older, are asked to participate in the control group of the study. For this purpose participating CCS are asked to contact all sisters meeting the inclusion criteria and to provide their contact information to the investigators. If a female survivor chooses to not register one or more available sisters, the reason is enquired about.

Approach of study participants and data collection

All eligible women receive a mailed package containing extensive study information, an informed consent and refusal form, and a questionnaire. They are asked to complete the questionnaire and return it with a signed informed consent form. Furthermore, they are asked to indicate on the informed consent form in which parts of the study they are willing to participate. In case of no response within 3 weeks, postal reminders are sent. When again after 3 weeks no response has been received, the women are contacted by telephone. Women who are not willing to participate in either part of the study are asked to complete a refusal form on which they can indicate the reason for not wanting to participate. These non-participants are asked to complete a brief questionnaire regarding parity, wish to have children, subfertility, subfertility treatment, and educational level in order to adjust for possible bias. The envelope containing the study information package is sealed and put in another envelop, together with a cover letter in which the study is explained very briefly. This is done in order to give the survivors the opportunity to choose whether or not they want to be confronted with the extensive study information. If not, they can send the unopened package return to sender. Figure 1 depicts the various response categories that apply to the DCOG LATER-VEVO study.
The data collected for survivors and siblings are the same, with the exception of data related to the anti-cancer treatment in the past. For both survivors and siblings information on reproductive and medical history is obtained by a questionnaire which is available either as hard copy or online. This questionnaire is an adaptation of a well-tested questionnaire used by the Department of Epidemiology of the Netherlands Cancer Institute in a large-scale Dutch cohort study of long-term effects of ovarian stimulation for in vitro fertilization [131 132]. It addresses the following issues: socio-demographic characteristics, menstrual history, desire to have children, reproductive history, pregnancy information, pregnancy outcomes, details of offspring, menopausal symptoms and menopause, use and duration of use of exogenous hormones, use and duration of use of fertility drugs and assisted reproductive techniques, family history of cancer and sub-/infertility, co-morbidities, and life style behaviour.

In order to assess reproductive function and ovarian reserve a blood sample is drawn and a transvaginal ultrasound of the reproductive organs is performed. From the blood sample, FSH, LH, estradiol, inhibin B, prolactin, and AMH concentrations are determined as well as the FSH receptor genotype. The ultrasound measurements, which assess the number of antral follicles in both ovaries as well as the length and width of the ovaries and the uterus, are performed by specifically trained personnel using a HD11 XE ultrasound system with a transvaginal probe which can perform three-dimensional (3D) imaging (EnVisor HD, Philips Medical Systems, Eindhoven, the Netherlands). First, a 2D ultrasound assessment of the pelvis is performed after which an automated mechanical sweep produces the 3D data. An ultrasound measurement is not performed when the participant indicates in the questionnaire that she has not yet been involved in sexual intercourse, unless she explicitly states she wants to undergo an ultrasound. Both the blood sampling and ultrasound measurements require specific timing. For both CCS and controls not using hormonal contraceptives this timing is as follows: (1) on day 2–5 of a natural menstrual cycle; (2) on any convenient moment in case of amenorrhea (no menses > 6 months). In those who are using hormonal contraceptives, alternative methods of timing were used (see section "The value of hormonal and ultrasound markers while using oral contraceptives").

Since January 2008, invitations have been sent to 1611 female CCS and 429 sister controls from all participating centers and data collection is still ongoing. As of March 1st 2012, 1215 CCS and 333 sister controls have responded, whereas from the remaining 396 survivors and 96 controls no response has been received to date. Table 1 describes the response and participation rates of CCS and controls as of March 1st 2012.

Table 1 Response and participation rates of childhood cancer survivors and sister controls in the DCOG LATER-VEVO study*

	Survivors	Sibling controls
Invited	1611	429
Response received (responders)	1215 (75%)	333 (78%)
No response received (non-responders)	396 (25%)	96 (22%)
Responders		
Participants		
Questionnaire only	306 (19%)	96 (22%)
Questionnaire and blood sample	126 (8%)	43 (10%)
Questionnaire, blood sample and transvaginal ultrasound	509 (32%)	174 (41%)
Non-participants	274 (17%)	20 (5%)

** Rates acquired as of March 1st 2012.*

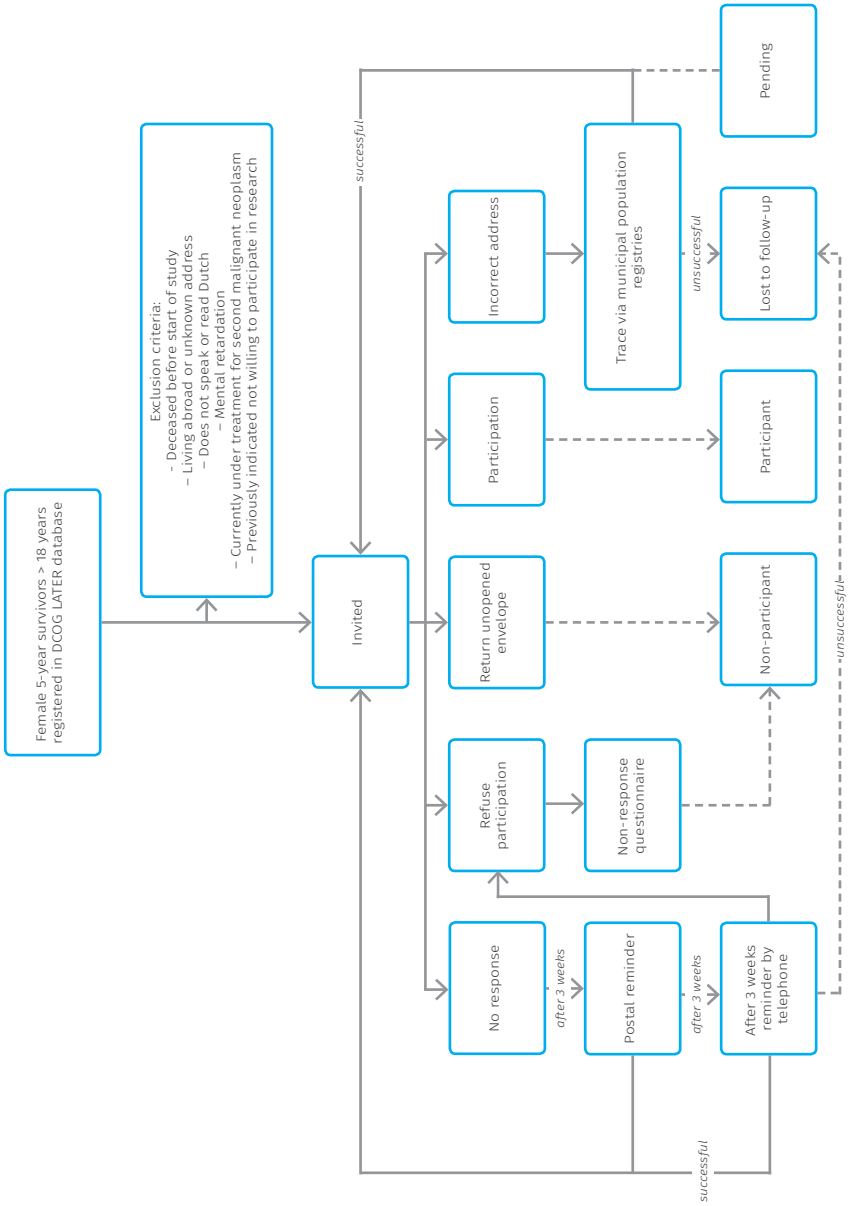


Figure 1 Response flow chart. The following categories and definitions are used to classify participants and non-participants in the DCOG LATER-VEVO study: 1) "eligible subjects" are individuals registered in the DCOG LATER database who were confirmed as meeting the study eligibility criteria; 2) "participants" are those who consented to participate; 3) "non-participants" are individuals who declined participation verbally or in writing, who returned the envelope with study information unopened or who, at first indicated they were willing to participate but ultimately did not do so; 4) individuals are considered "lost to follow-up" if they were not located after intensive tracing efforts; and 5) individuals are classified as "pending" when they are actively being traced and recruited to participate.

CHALLENGES

The challenges encountered during the study can be divided into two categories: challenges related to the study population and challenges related to the data collection procedures.

Study population

Bias due to different characteristics of the women in the various response groups

Eligible study subjects can either respond (responders) or not respond to the study invitation (non-responders). The responders either decide to participate (participants) or not to participate in the study (non-participants). When they decide to participate they can subsequently choose to take part in one, two or all three parts of the study. In total, three groups of participants are distinguished: 1. questionnaire only; 2. questionnaire and blood sample; 3. questionnaire, blood sample and transvaginal ultrasound. The responders may not be comparable to the non-responders and the same is true for non-participants versus participants. In addition, subjects who choose to complete the questionnaire only may not be comparable to those who also participate in the clinical assessment. This may potentially lead to selection bias, which may influence the validity of the study results. Therefore, it is important to identify potential differences between the various response groups in order to be able to control for selection bias during the data analyses of the DCOG LATER-VEVO study.
In order to identify the presence and direction of possible selection bias interim data analyses were performed, in which differences between the characteristics of responders and non-responders were compared. Results showed that the age at study invitation (28.8 vs. 28.2 years), age at diagnosis (6.96 vs. 7.06 years) and time since treatment (7998 vs. 7658 days) were not different between responders and non-responders. In addition, differences between the characteristics of participants and non-participants were compared. Results showed that the non-participating CCS ($n = 274$) did not significantly differ from the participating CCS regarding current age ($p = 0.09$). In addition, there were no significant differences in age at diagnosis or in time since diagnosis ($p = 0.23$ and $p = 0.24$, respectively). Of the 274 non-participants, 17% ($n = 46$) were willing to complete the brief non-participant questionnaire. Data from this questionnaire showed no significant differences with regard to educational level, although the proportion of women with a high educational level in the participants group was substantially higher compared to the non-participants group (39.9% vs. 26.1%, $p = 0.09$). Moreover, a larger proportion of the non-participants reported to already have offspring in comparison with participants. However, this difference was not significant either (46.3% vs. 33.9%, $p = 0.10$). Nevertheless, this may suggest that women with proven fertility may be less likely to participate in the study than those who have not given birth (yet). This implies that caution should be taken when interpreting the results of the DCOG LATER-VEVO study since an overestimation of the adverse effects of the cancer treatment on reproductive outcomes (i.e. actual fertility) might be introduced. We realize, however, that the number of women completing the non-participant questionnaire was rather low.

As a consequence, this group of women may not be fully representative of the total group of non-participants. However, the non-participants that did and those who did not complete the non-participant questionnaire appeared not to significantly differ regarding current age, age at diagnosis, and time since diagnosis.
With regard to possible selection bias within the participating groups, we evaluated the differences in characteristics between those participating in the questionnaire part only and those participating in both the questionnaire and the clinical part (blood and/or ultrasound) of the study (Table 2). It appeared that women who completed the questionnaire only were older, had a longer follow-up time since diagnosis, were less likely to be highly educated, were more likely to have had intercourse and to have offspring than women who also agreed to participate in the clinical part.

Table 2 Comparison of several characteristics between two different groups of participants

Variable	Participants		P value
	Participating in questionnaire part only (n=306)	Participating in questionnaire and clinical part (n=635)	
Age (years)	30 (18-52)	27 (18-56)	< 0.001
Age at diagnosis (years)	6 (0-16)	6 (0-17)	NS
Time since diagnosis (days)	8281 (2563-15423)	7418 (2068-14612)	0.001
High education (n/N) (%)*	103/300 (34.3)	267/619 (43.1)	0.002
In committed relationship (n/N) (%)*	217/304 (71.4)	431/633 (68.1)	NS
Has had intercourse (n/N) (%)*	248/303 (81.8)	555/631 (88.0)	0.01
Has offspring (n/N) (%)*	146/304 (48.0)	183/633 (28.9)	< 0.001
Has previously consulted a gynecologist for fertility problems (n/N) (%)*	46/289 (15.9)	72/621 (11.6)	NS

*Data presented as median (range) unless indicated otherwise. NS = not significant. * Total N does not correspond with the number mentioned in the heading of the table because of missing data.*

Participants who ultimately did not show up for outpatient clinic visit

If participants decide to provide a blood sample and/or undergo a transvaginal ultrasound, they are asked to contact the research staff on a specific day of their menstrual cycle in order to plan the clinical assessment at the outpatient clinic. However, we experienced that some participants who initially consented to a clinical assessment did not (yet) follow-up on their decision 6 months after their written consent (n = 90). Twenty-seven women (30%) have expressed a reason (pregnancy, breastfeeding, illness or disability, and personal or family responsibilities) while the

remaining 63 have not yet contacted the research staff, even after several reminders. Ultimately, these women will be classified as non-participants to the clinical part of the study.

Finding an adequately sized and well-matched control group

Sisters of participating CCS are invited as controls, since they have the same genetic and socio-economic background (which might influence fertility and other outcomes). However, the inclusion of only sisters in the control group has shown to result in insufficient numbers compared to the number of CCS (941 versus 313, see also Table 1). The reason is two-fold: not all participating CCS have an eligible sister aged 18 years or older and not all CCS with eligible sisters gave permission to contact their sisters for the control group. For the DCOG LATER-VEVO study this may ultimately result in insufficient power for certain subgroup analyses. Moreover, not all research questions of the study require a control group that is genetically and/or socioeconomically comparable to the survivor group. Therefore, it was decided to expand the control group by including women from the general population as well.
For this purpose, general practitioners of the participating CCS are asked to randomly select and invite subjects from their female patient population. For logistical reasons, we selected general practitioners practices located in the area surrounding the coordinating center. In order to ensure a comparable age distribution between CCS and controls, these general practitioners are asked to select women within a specified age range of five years (so-called "GP controls"). For these women, the same inclusion criteria apply as for the sister controls: never to have been diagnosed with cancer, able to read and speak Dutch, and 18 years or older. This method of approach resulted in 1184 women who were invited to participate in the DCOG LATER-VEVO study as controls. So far, 935 (79%) have responded and 429 women have consented to participate, 308 of whom in the clinical part of the study.
Within the group of eligible controls recruited via the general practitioners, we conducted a non-responder analysis. A random sample (n = 200) was drawn from the GP controls who did not respond to the study invitation. From these women the following variables were collected from the medical records at the general practitioner's office: age, having offspring, maternal age at first childbirth, fertility-related problems, and visits to gynecologists. These data, which were made anonymous for privacy reasons, were compared with data of a random sample of GP controls who did respond to the study invitation (n = 194). Preliminary results show no significant differences between the two groups with respect to the before-mentioned variables. This suggests that the degree of selective participation within the GP control group is low.
We also evaluated whether sister controls differed from the GP controls on several basic characteristics (Table 3). Results show that GP controls were more likely to have a high educational level. In addition, these women were older than sister controls. There were no significant differences between siblings and GP controls with regard to relationships, offspring or fertility issues.

Table 3 Differences between sister controls and controls recruited through the general practitioner (GP controls)

	Sibling controls (n=313)	GP controls (n=430)	P value
Age	30 (18-58)	34 (18-54)	< 0.001
High education (n/N) (%)	158/308 (51.3)	293/428 (68.5)	< 0.001
Is in committed relationship (n/N) (%)	249/311 (80.1)	341/428 (79.7)	NS
Has had intercourse (n/N) (%)	288/312 (92.3)	407/426 (95.5)	NS
Has offspring (n/N) (%)	140/313 (44.7)	201/429 (46.9)	NS
Has previously consulted a gynecologist for fertility problems (n/N) (%)	29/301 (9.6)	44/411 (10.7)	NS

Data presented as median (range) unless indicated otherwise. NS= not significant.

Data collection

Differences in response between web-based and paper-based questionnaires

Web-based questionnaires have several advantages over paper-based questionnaires. They are less time-consuming, less costly, and the data of the respondents are already available in an electronic format, leading to less input errors [133-135]. At the time the DCOG LATER-VEVO study was set up, no literature was available concerning the differences in response rates of CCS to a web-based or a hard copy paper questionnaire, when they are offered both types of questionnaire. Therefore, we conducted a nested randomized study to evaluate whether the use of either the web-based or hard copy paper questionnaire resulted in differences in response rates, type of response, and characteristics of the (non-)responders. In this study, 277 eligible women were randomly selected to receive either a mixed invitation (a hard copy paper questionnaire together with the login details for the web-based questionnaire) or a web-only invitation (login details only). Women receiving the web-only invitation were given the opportunity to apply for a hard copy of the questionnaire by returning a form. The results showed that although the overall response rates to both types of invitation were similar, adding a paper version of a questionnaire to a web-only invitation resulted in more respondents completing the hard copy of the questionnaire. In addition, women who were older, higher educated as well as those who were a student, had a higher probability of completing the web-based version of the questionnaire (Table 4) [136]. It was decided that future invitations for the DCOG LATER-VEVO study should include both a hard copy of the questionnaire and the login details for completing the web-based questionnaire.

Validity of self-reported outcomes

Previous studies on fertility and pregnancy outcomes in CCS were often designed as large cohort studies in which the outcomes of interest were obtained through interviews or mailed questionnaires [117 124]. However, the reliability and validity of self-reported data can be limited and one should be aware of possible (non) differential misclassification bias. Several studies among healthy women have

Table 4 Factors associated with the probability of completing the web-based version of the questionnaire: results of logistic regression*

	OR (95% CI)
Age	1.08 (1.02-1.15)
Educational level (ref. group: High level)	
Medium	0.65 (0.28-1.53)
Low	0.06 (0.01-0.52)
Employment status (ref. group: Employed)	
Student	3.25 (1.00-10.56)
Unemployed	0.35 (0.10-1.29)
Randomization group (ref. group: Mixed invitation group)	2.85 (1.31-6.21)

OR, odds ratio; CI, confidence interval.

** Derived from Van den Berg MH et al. Using web-based and paper-based questionnaires for collecting data on fertility issues among female cancer survivors: differences in response characteristics. J Med Internet Res. 2011 Sep 29;13(3):e76.*

been performed assessing the accuracy and validity of self-report for pregnancy outcomes. Data from these studies showed that birth weight and gestational age were accurately reported, but time to pregnancy, reasons for subfertility, and maternal and neonatal complications during labour and delivery were reported with less accuracy [137-139]. Since no literature was available assessing the validity of self-reported pregnancy outcomes by CCS, we conducted a validation study among our study participants [140]. Women were eligible for the validation study when they reported in the questionnaire to have had a child between 1/1/1985 and 31/12/2009. Reference data on pregnancies and pregnancy outcomes were abstracted from the Netherlands Perinatal Registry (PRN). In this nationwide population-based registry midwives, obstetricians and pediatricians register pregnancy outcomes of all births and data are available from 1985 to 2009. Records of self-reported pregnancies were linked to the PRN by using both the mother's date of birth and the child's date of birth as linkage keys. At the time of the validation study, 879 CCS and 287 controls had returned the study questionnaire. In total, 589 pregnancies were reported in 289 CCS compared to 293 pregnancies in 123 controls. Linkage to the PRN yielded 510 unique hits (345 pregnancies in 186 CCS, and 186 pregnancies in 87 controls). A high intra-class correlation coefficient (ICC) was found for birth weight (0.94 (95%CI 0.91-0.96) and 0.87 (95%CI 0.83-0.90) for CCS and controls, respectively). For gestational age, the ICC was 0.88 for CCS (95%CI 0.85-0.91), but only 0.49 for controls (95%CI 0.32-0.62). The kappa value for method of conception was moderate to good, but varied largely per method (0.56 for hormonal stimulation to 1.0 for IUI). The kappa values for different methods of delivery were good for CCS and controls (0.76 for spontaneous delivery to 0.92 for vacuum/forcipal extraction). Kappa for pregnancy-induced hypertension was 0.59 for CCS and 0.61 for controls. Multilevel analyses showed no differences in accuracy associated with time since pregnancy or educational level.

The value of hormonal and ultrasound markers while using oral contraceptives

The results of a pilot study conducted before the start of the DCOG LATER-VEVO study showed that 55% of the CCS used oral contraceptives. As it was anticipated that the response rates to the study would be significantly lower when the participants had to stop using oral contraceptives to be eligible for our study, we conducted a study to compare both hormonal and ultrasound markers of ovarian reserve measured on day 7 of the pill free interval and two subsequent natural cycles. Results showed that FSH and inhibin B values decreased significantly when contraception use was discontinued, whereas values of AMH, AFC and ovarian volume increased significantly. Thus, hormonal and ultrasound markers of ovarian function in oral contraceptive users measured at the end of the hormone-free interval do not fully represent subsequent natural early follicular phase values. However, FSH, AMH and AFC can be used to predict early follicular phase values using calculated prediction equations [111].
The results of this study have led to the following procedures regarding the timing of the clinical measurements of the DCOG LATER-VEVO study. Women who are on oral contraceptives or who use a combined contraceptive vaginal ring are asked to refrain from using these and use other methods of birth control (condoms were provided free of charge), for at least two months prior to the outpatient clinic visit. Depending on the menstrual pattern they develop, these participants are invited according to the timing schedule described above (see section “Approach of study participants and data collection”) during the second natural menstrual cycle. Women who do not wish to discontinue the use of oral contraceptives or vaginal ring during the study period are invited on day 7 of the pill-free or ring-free period for the clinical measurements. Women who use a hormone-containing intrauterine device (IUD) are asked to monitor their basal body temperature (BBT) daily for at least 4 weeks. The BBT chart was evaluated by an experienced gynecologist (CBL) to detect ovulation. Ovulation is confirmed when the chart appeared to be biphasic (temperature shift of at least 0.5 °C). The date of the assessment for the study is then planned in the early-follicular phase. A monophasic BBT chart was deemed anovulatory and the measurements are planned at any convenient moment. Women using long-acting contraceptive injections or women with a contraceptive implant are excluded from the clinical part of the study.

Blinding and inter- and intra-observer variation of ultrasound measurements

To assess ovarian function and ovarian reserve all study participants who consent to this procedure undergo a transvaginal ultrasound. This measurement is performed in five centers across the Netherlands, making this examination as convenient and as timesaving for the participant as possible. All centers are equipped with the same type of 3D ultrasound apparatus. It is not possible to blind ultrasonographers to the CCS status of the participants given the fact that the participants themselves are evidently not blinded to their status and often ask questions regarding their prior cancer treatment during the clinical visit. However, the ultrasonographers do not have access to diagnostic or treatment data prior to or during the procedure.

Moreover, the stored 3D files are made anonymous and therefore, no prior knowledge regarding the diagnosis or treatment is available to the investigator analysing the 3D data. This will minimize observer bias.
3D ultrasound is capable of visualizing all three orthogonal planes simultaneously. With the stored volumetric data, imaging can be accurately evaluated offline. Literature shows that both 2D and 3D ultrasound measurements have high intra-observer and inter-observer reliability [141-144], but some limitations were found in the between-method reliability and the degree of agreement when higher numbers of follicles were counted [144]. However, these validation studies have all been performed in healthy controls, or in women undergoing IVF/ICSI treatment. The 2D and 3D techniques have not been validated for women treated for childhood cancer in the past. Former treatment might have induced changes to the reproductive organs, which might result in changes in the intra- and inter-observer reliability of both methods. Moreover, the between-method reliability of real-time 2D images and stored 3D images acquired from CCS has not been investigated so far. Therefore, we are conducting an evaluation of the intra-observer, the inter-observer, and the between-method reliability of both the 2D and 3D ultrasound measurements of the DCOG LATER-VEVO study.

DISCUSSION

This study on reproductive function, ovarian reserve, and risk of premature menopause in female childhood cancer survivors is the first large nationwide late effects study in the Netherlands. Compared to previously conducted studies on reproductive outcomes in CCS, this DCOG LATER-VEVO study has several strengths.
First, the results of the study are not solely based on self-reported data from questionnaires. Clinical data on ovarian and uterine function are included as well, which allow for a more objective evaluation of the actual fertility status. The extensive set of data acquired in this study will result in detailed knowledge regarding treatment-induced effects on the female reproductive potential, particularly the effects of different types of treatment, doses of drugs, radiation sites and doses, and age at time of treatment.
Second, when data collection of the study is finalized, the reproductive data of the CCS can be compared with those of a large number of controls. For these controls questionnaire as well as clinical data are available. The size of the control group as well as the availability of clinical data from these controls can be considered unique study features within the field of late effect studies among CCS.
Third, the complete cohort of adult female childhood cancer survivors, treated in one of seven Dutch pediatric oncology centers, is invited to participate in the study, thereby minimizing the risk of selection bias due to loss to follow-up. This is possible because recently a database containing up-to-date patient and treatment data of all Dutch 5-year CCS diagnosed before 2002 has been established. In addition, through the Dutch system of municipal population registries, which fully cover the Dutch population, individuals can nearly always be traced, despite frequent moving. Furthermore, inclusion of the complete cohort provides the advantage of a well-powered study in which several subgroup analyses can reliably be performed.

Table 5. Challenges and recommendations

	Challenge	Recommendations
Study population	Dealing with participation bias • Responders and non-responders • Participants and non-participants • Different types of participants • Participants lost to follow-up for the clinical assessment	• Keep non-response or loss to follow-up to a minimum • Characterize non-responders or those lost to follow-up • Control for extent and direction of bias in final data analysis
	Finding an adequately sized and well-matched control group	• In case the number of controls is insufficient: incorporate other types of control subjects • Choose types of controls that are representative of the study population • Characterize and control for differences between survivors and controls
Data collection	Validating instruments for data collection	• Compare self-reported data with an more objective source, such as medical records or registries • Conduct reliability studies to account for inter- and intra- observer variation • If possible, use data collection instruments that allow for one investigator to analyse collected data (observer bias)

Despite the above-mentioned strengths of the DCOG LATER-VEVO study, we have also identified several challenges in the design and conduct of the study. To address these challenges, several recommendations have been formulated (Table 5).
Our results have shown that there are no significant differences between participants and non-participants regarding socio-demographic data. Nevertheless, a trend was seen towards more highly educated women in the participant group, as well as lower offspring rates. Differences between participants and non-participants may introduce participation bias. This type of bias can be minimized when the number of subjects refusing to participate is kept to a minimum. Moreover, it is important to ask women who do not wish to participate to complete a brief non-participants questionnaire in order to be able to characterize this group. By the time data collection of the DCOG LATER-VEVO study is completed, characteristics of participants and non-participants will be compared once again. By doing so, the extent and direction of participation bias can be established, which can be taken into account when analyzing the data.
During the study period it became clear that a group of participants, who initially consented to visit the outpatient clinic for blood sampling and/or ultrasound measurement, ultimately did not do so. We are making every effort to keep the number of these 'no-shows' as low as possible since this selective non-participation might add to selection bias. Indeed, those who feel less inclined to visit the clinic might differ from those who do [145]. At this time, the direction and the magnitude of this bias is difficult to predict. It might be the case that those who ultimately do not visit the clinic may be more fertile and do not see the need for the clinical measurements (anymore). However, these women might also have obtained

information that they are infertile. It is therefore of paramount importance that basic information regarding the main outcomes of this study is also collected for the women who ultimately did not show up for the clinical visit. Fortunately, many of these women are seen periodically in outpatient clinics throughout the country for late effects follow-up screening. This enables us to retrieve information regarding reproductive outcomes of these women and to further investigate possible bias.

Two types of controls are included in the study, i.e. sisters of participating CCS and women from the general population recruited through general practitioners. For our study, siblings form a better control group than a random sample from the population, based on a municipal registry sample, neighbourhood controls, or friends and relatives. They share the same genetic and socio-economic background as the CCS. Furthermore, siblings may be more motivated to participate for altruistic reasons, i.e. participating in the study for the sake of their sister, whereas controls from a random sample in the population might have other reasons to participate, for example, fertility problems. This could also lead to selection bias, which could consequently influence our study results. The strong recommendation to register all sisters, and not to choose only one, may further reduce this form of sampling bias.

However, when including only sisters as controls, the number of controls would be significantly lower than the number of survivors. In order to attain sufficient power for the planned statistical analyses, we chose to also include women from the general population. By using a 'targeted' invitation strategy we aimed to minimize the risk of selection bias caused by selective participation of, for example, women experiencing fertility problems. Several general practitioners were asked to randomly select and invite women with a specific year of birth from their patient population. In this way, bias due to selective participation is expected to be less compared to applying a broader invitation strategy, for example by using advertisements in daily newspapers to recruit controls. However, it is still probable that some form of selection bias has occurred, since comparisons between the sister controls and GP controls show that the latter group is older and more likely to have a higher educational level.

In order to minimize the travelling time associated with attending the outpatient clinic, only general practitioners located close to the coordinating center (VU University Medical Center Amsterdam) were approached for the recruitment of controls. We realize that this method may introduce additional bias, because of regional differences, which can ultimately lead to selective participation.

Another limitation of the study is the fact that CCS and controls may differ regarding their ability to accurately recall information about past (medical) events, leading to so-called recall or misclassification bias. CCS might better recall their past medical history, because they are frequently seen in outpatient clinics for follow-up, whereas controls may have forgotten about less severe medical problems they have encountered in the past. This might lead to an overestimation of the risk of medical problems in the survivor group. Furthermore, most CCS are aware of the fact that their previous cancer treatment may have had detrimental effects on their fertility. As a consequence, they may report more accurately on several fertility-related items since they have been (or still are) more 'focused' on fertility-related events from the past. There is no method available to correct for this type of bias. Nevertheless,

the use of objective measurements, such as hormonal analyses and ultrasound examinations, and the validation of self-reported data by comparing with medical records or available registries may minimize the influence of this type of bias.
In order to evaluate the validity of the several outcomes in the study questionnaire, we conducted a study in which self-reported data on pregnancies were compared to data from a nationwide registry. Overall, self-reported pregnancy outcomes of CCS indeed appeared to agree better with the registry data than those of controls. This might be due to the increased awareness of late effects and a higher frequency of medical follow-up. Although self-reported data regarding fertility and pregnancy by CCS seem consistent with registry parameters, differential misclassification between CCS and controls may occur and should be taken into account when interpreting the data. In the near future, we also aim to validate data on fertility issues and gynaecological disorders by comparing the questionnaire data to the medical records.
Clinical measurements for the study were performed at several locations throughout the Netherlands. In order to rule out observer bias, all ultrasound data will be analysed by one investigator using stored 3D ultrasound data. In addition, all endocrinological measurements will be done in a single laboratory at the end of the study, as it is not unlikely that changes in laboratory kits will occur during the study period.
In conclusion, the results of this nationwide study will provide valuable information for counselling female childhood cancer patients and survivors regarding their reproductive potential now and in the future. Other investigators planning to conduct large nationwide cohort studies on late effects may encounter similar challenges as those encountered during this study. Our experiences as well as the way we addressed these challenges will hopefully contribute to an optimal design and conduct of future late effects studies.

516278-L-bw-Overbeek
Processed on: 19-1-2018

4

Fertility studies in female childhood cancer survivors: selecting appropriate comparison groups

M.H. van den Berg
E. van Dulmen-den Broeder
A. Overbeek
C.M. Ronckers
W. van Dorp
L.C. Kremer
M.M. van den Heuvel-Eibrink
G.A. Huizinga
J.J. Loonen
A.B. Versluys
D. Bresters
C.B. Lambalk
G.J.L. Kaspers
F.E. van Leeuwen

Reprod Biomed Online.
2014 Sep;29(3):352-61.

ABSTRACT

Little information is available on the use of appropriate comparison groups for studies investigating late effects of childhood cancer. We report on our experiences regarding the recruitment of subjects for two comparison groups in a nationwide study on reproductive function and ovarian reserve in female childhood cancer survivors (the DCOG LATER-VEVO Study). Two types of comparison groups were used, i.e. sisters of participating survivors and controls from the general population (GP). Sixty-one percent (352/580) of the participating survivors who had a sister gave permission to invite them for the study. The participation rate of sisters was much higher than of GP controls (74% vs. 21%, respectively), whereas considerably more effort was involved in recruiting GP controls. GP controls were significantly older and higher educated than sister controls (both P values < 0.001). However, no significant differences between both types of comparison groups existed regarding several fertility-related characteristics, suggesting minimal bias due to selective participation. Researchers setting up a study to investigate late effects among survivors of childhood cancer, should carefully consider the advantages and disadvantages of using various types of comparison groups.

INTRODUCTION

Over the past 40 years, advances in diagnosis and treatment have substantially improved survival of most childhood and adolescent cancers, resulting in overall 5-year survival rates of over 80% in Europe [146], as well as in the USA [147]. As a result, there is a rapidly growing population of young adult survivors of childhood cancer. However, cancer treatment during childhood can induce complications which may not become apparent until many years later. These treatment-related late effects include, among others, secondary neoplasms, cardiac dysfunction, reduced growth, sub- or infertility, impaired cognitive function, psychosocial problems, and a reduced quality of life [36 37 148].

In the past decades many studies have been conducted to assess the long-term adverse effects of treatment in childhood cancer survivors (CCSs) or to evaluate screening and surveillance programs for this group of patients. Study designs typically included retrospective cohort studies, (nested) case-control studies or cross-sectional studies [149 150]. Moreover, the occurrence of adverse events among CCSs has been contrasted to different comparison groups. Typical comparison groups that have been used in late effect studies include subjects who were not treated for cancer during childhood (e.g. siblings or individuals recruited from the general population) or CCSs who were not exposed to a certain type of treatment. In addition, data from CCSs have been compared with readily available (inter)national population norms.

Several reports of late effect studies have extensively described the study design, as well as the procedures for participant recruitment and data collection [151-156]. Also, methodological issues or pitfalls regarding these studies have been summarized [149 150 157-161]. A frequently reported limitation of late effect studies is the lack of appropriate comparison groups [158 160]. However, to date, no published reports comprehensively summarize the advantages, disadvantages and experiences regarding the recruitment of different types of comparison groups, while deciding on the type, size, and number of comparison groups, as well as on the procedure for recruiting the control subjects is an important and difficult task when setting up a study.

In the Netherlands the so-called DCOG LATER-VEVO study is being conducted, which is a nationwide study on reproductive function, ovarian reserve, and premature menopause in female CCSs [162]. Since in this study several types of comparison groups were used, the study provides an ideal opportunity to reflect on which comparison group seem to be most appropriate when conducting such a study. In this report we therefore aim to describe the advantages and disadvantages of using different types of comparison groups by reporting on our experiences regarding the recruitment of subjects for comparison groups for the DCOG LATER-VEVO study.

METHODS

Design of the DCOG LATER-VEVO study

The DCOG LATER-VEVO study was initiated in the Netherlands in 2006 as a nationwide retrospective cohort study evaluating the effects of cancer treatment on reproductive function, ovarian reserve and risk of premature menopause in female CCSs. The study design and cohort characteristics have been described in a previous report [162]. In short, the study consists of three parts: (1) a questionnaire; (2) blood sampling for serum hormone levels; and (3) a transvaginal ultrasound measurement of the reproductive organs. We aim to invite all eligible subjects from a cohort of 5-year survivors treated for childhood cancer between 1963 and 2002 at one of the seven Dutch pediatric oncology - and stem cell transplant centers, collectively known as the Dutch Childhood Oncology Group - Long-term Effects after Childhood Cancer cohort (DCOG LATER cohort). For the DCOG LATER-VEVO study, the eligible cohort consists of 1860 female CCSSs. Data collection is still ongoing. The study was approved by the Medical Ethical Committee of VU University Medical Center Amsterdam (reference No. 2006/249; approved January 4, 2007), and informed consent was obtained from all participants.

Comparison groups in the study

Initially, only sisters of participating CCSs were invited to participate in the comparison group of the DCOG LATER-VEVO study. For the main outcomes of this study sisters were primarily considered the most appropriate comparison group as they broadly share the same genetic and socio-demographic background, variables that might influence fertility and related outcomes. However, not all research questions or outcomes of the DCOG LATER-VEVO study require a comparison group that is genetically and/or socio-demographically comparable. Moreover, for some outcomes siblings may in fact not be the most optimal comparison group as the experience of childhood cancer may not only impact future decisions concerning fertility or the psychosocial state of CCSs but also of siblings [163 164]. Furthermore, when restricting the comparison group to sisters only, the number of included controls would be much lower than the number of female CCSs, since not all participating CCSs had an eligible sister and not all CCSs with eligible sisters gave permission for sibling contact. Therefore, to ensure sufficient power for all intended (subgroup) analyses of the DCOG LATER-VEVO study, we decided to expand the comparison group by also including age-matched women from the general population.

Sister comparison group

Recruitment from female CCSs

Sisters of CCSs were recruited by asking all participating CCSs if they had any sisters aged 18 years or older. Subsequently, those who had at least one eligible sister were asked for permission to contact and invite their sister(s) for participation in the comparison group of the study. We purposefully asked the CCSs to register all eligible sisters in order to prevent selection of a specific sister into the study.

However, in case a CCS did not want to register all sisters, we offered the possibility of registering just one. For survivors not allowing registration of their respective siblings, we investigated the underlying reasons in two different ways: 1. a multiple-choice question in the questionnaire inquired about the main reason for not registering (all) eligible sisters; 2. in 2010 a qualitative study was conducted among a random subgroup of CCSs in which the reasons for not registering a sister were further examined by means of a semi-structured interview.

Recruitment from male CCSs

During the first year of the DCOG LATER-VEVO study it became clear that the number of sisters of female CCSs serving as controls in the study would be less than expected. Therefore, we pilot-tested the recruitment of sisters from male CCSs into the study at the VU University Medical Center Amsterdam (VUmc). The method of recruitment was identical to the method used among female CCSs (see above). Obviously, a disadvantage of using sisters of non-participating male CCSs as controls in the study, is the fact that these controls (as opposed to sisters of participating female CCSs) are not genetically nor socio-demographically linked to the participating female CCSs. Nevertheless, we expected that the recruitment of siblings, either from female or male CCSs, would be less prone to selection bias from non-participation.

Women from the general population (GP)

General practitioners of CCSs participating in the DCOG LATER-VEVO study received a letter in which they were asked to help recruiting women for the comparison group. For logistic reasons, only general practitioner practices located in the area surrounding the coordinating center (Amsterdam) were asked to cooperate. In case the general practitioner was willing to cooperate, we asked him/her to select all women from their patient population who had the same year of birth (plus or minus two years) as the CCS in their patient population. By using this method of frequency matching, we aimed to acquire similar distributions over age strata within the various study groups [390]. All eligible women were then sent a letter, signed by the general practitioner, in which the study was briefly explained. A response form and a pre-stamped envelope accompanied this letter. On this form women could indicate whether they were interested in the study and wanted to receive the extensive study information package. Women who received the extensive study invitation and did not respond within 3 weeks, were sent a postal reminder, and, in case again no response was received in the following 3 weeks, were contacted by telephone. Similar to sisters, the GP controls obtained through the general practitioner could choose to participate in the questionnaire part, the clinical part or both parts of the study. All women who participated in the clinical part of the study (i.e. blood and ultrasound measurement) and who indicated that they wanted to receive their personal results, were sent this information afterwards.

Data analysis

Comparisons between CCSs, sisters and GP controls were analyzed using SPSS for Windows-version 20.0 (SPSS Inc., Chicago, Illinois, USA). Differences between the groups were analyzed by the Mann-Whitney U test or chi-squared test where appropriate. Furthermore, multivariable logistic regression analysis was used to compute age-adjusted p-values regarding these differences as well as to investigate which factors are associated with the probability of giving permission to approach eligible sisters by CCSs (odds ratios and 95% confidence intervals).

RESULTS

Sister controls recruited through female and male CCSs

Questionnaire data showed that 55% (n=580) of the participating female CCSs (n=1051) had one or more sisters who were potentially eligible for the study (799 sisters in total). However, only 352 (61%) out of the 580 participating CCSs gave permission to approach one or more of their sisters. It appeared that most gave permission to approach either *all* (56%) or *none* of their eligible sisters (39%). In total, 447 sisters received a study invitation. In the pilot study to recruit sisters from male CCSs at VUmc 97 male CCSs were approached. In total, 49 appeared to have one or more eligible sisters and 30 (61%) gave permission to approach their sister(s), a proportion similar to the female CCSs (see above). Ultimately, we obtained permission to approach 484 sisters in total: 447 through female and 37 through male CCSs.

Reasons of CCSs for not giving permission to approach sisters

Of the remaining 228 out of 580 female CCSs, who had one or more eligible sisters but chose not to register any of them, 144 (63%) provided reasons for not doing so in the questionnaire. The main reason was that CCSs did not want to burden their sisters with the study (47%). In addition, 11% indicated that they were already aware of the fact that their sister was not interested in the study, and again 11% reported their sister to have personal problems (not being fertility-related) at time of the study. In addition, 10% of CCSs did not want to confront their sisters with the past. The remaining 21% include various other reasons.

Of the 55 randomly chosen CCSs who were asked to undergo a brief semi-structured telephone interview, 44 agreed to participate. Results show that the majority (n=26/44, 59%) had more than one reason for not allowing us to approach their eligible sisters. It appeared that the reasons mentioned by CCSs mainly reflected their own opinion since most (n=22/37, 59%) said they had not discussed potential study participation with their sisters. When CCSs had previously reported that they considered the burden for their sisters as high, they meant both a practical burden (e.g. time constraints) as well as a mental burden (e.g. too confronting) in most cases. In addition, a small group (n=9) indicated that the importance of their sister participating in the study had not been clear to them. Moreover, it appeared that some CCSs (n=4) did not give permission to approach their sister(s) because they

wrongly assumed that their sister(s) was/were not eligible for the study. All in all, these results indicate that the reason for not allowing us to approach eligible sisters does not appear to be fertility-related.
Furthermore, results showed that the probability of giving permission to approach sisters was significantly higher for CCSs who: 1. had a high or medium educational level (vs. a low educational level) (P = 0.01); 2. were divorced or married (vs. never married) (P =0.03); 3. were a student or gainfully employed (vs. unemployed) (P = 0.001); 4. were participants in both the questionnaire and the clinical part of the study (vs. participants in the questionnaire part only) (P < 0.001). Age did not appear to influence the probability of registering sisters (Table 1).

Table 1 Factors in female childhood cancer survivors associated with the probability of giving permission to approach eligible sisters

	P value	OR	95% CI	
			Lower	Upper
Educational level (ref. group: Low level)*	0.01			
Medium		1.71	0.78	3.72
High		3.03	1.33	6.90
Marital status (ref. group: Never married)	0.03			
Married		1.80	1.12	2.89
Divorced		3.58	0.64	20.03
Employment status (ref. group: Unemployed)	0.001			
Employed		2.02	1.09	3.77
Student		4.38	1.98	9.67
Study parts (ref. group: Questionnaire only)	< 0.001			
Questionnaire and blood sample		1.86	0.97	3.57
Questionnaire, blood sample, and ultrasound		3.16	2.04	4.87

** Low educational level: up to and including lower technical and vocational training; medium: up to and including secondary technical and vocational training; high: up to and including higher technical and vocational training and university. OR = odds ratio, CI = confidence interval.*

Participation rates of sister controls

Figure 1 shows the response rate of all sisters who were sent a study invitation (n=484). The proportion of non-responders was relatively low (21%). However, among the responders group, there was a small proportion (5%) of sisters who actively indicated that they were not willing to participate. Overall this resulted in a participation rate of 74%. Eventually, more than one third (35%) of the invited sisters participated in all three study parts, while 7% completed the questionnaire and provided a blood sample and 31% completed the questionnaire-only.

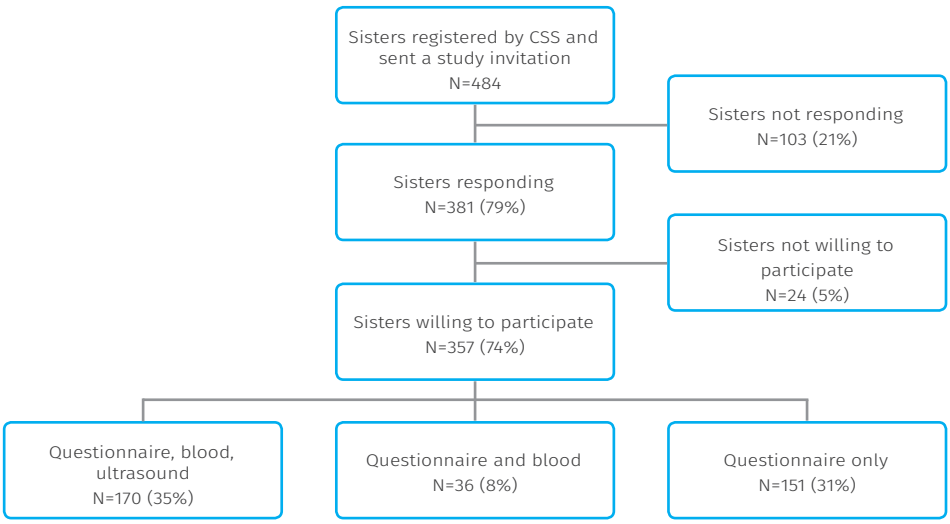


Figure 1 Response of registered sister controls. All response and participation rates are calculated using the initial number of registered sisters (n=484) as the denominator.

Participation rates of GP controls

Fifty general practitioners located in the area of Amsterdam were requested to aid in recruiting controls for the DCOG LATER-VEVO study and 41 (82%) general practitioners responded. However, considerable effort had to be made to acquire this high response: 84% (42/50) of the general practitioners was sent a reminder letter by post and 34 (81%) of those not responding to the postal reminder had to be contacted by telephone in order to eventually get a response.

Eventually, 19 (38%) general practitioners were willing to cooperate. They sent the short study invitation letter to 2,120 women from their patient populations. Figure 2 shows the number of women responding to this letter, the number of women receiving the extensive information package, as well as the number of women eventually participating in the DCOG LATER-VEVO study.

It appeared that of the 1,203 women (57%) who responded to the invitation letter, 727 women (60%) were interested in the study and were sent the extensive study invitation. Results show that the response rate to this invitation was 68% (493 women), with the majority of the women willing to participate (90%). However, the ultimate participation rate (i.e. the number of women eventually participating compared to the number of women who were sent the initial short study invitation letter) was rather low (21%).

In order to get some insight into the possible introduction of bias due to non-participation into the study we evaluated to which extent the GP controls who participated in the DCOG LATER-VEVO study differed from those who decided not to participate or those who did not respond at all. For practical reasons this study was performed in a single general practitioner's office. Of the 992 invitations that were sent out by this office, 194 women agreed to participate in the DCOG LATER-VEVO study (the "participants"), while 171 women replied that they were not willing to participate (the "refusers"), and 627 women did not respond to the invitation

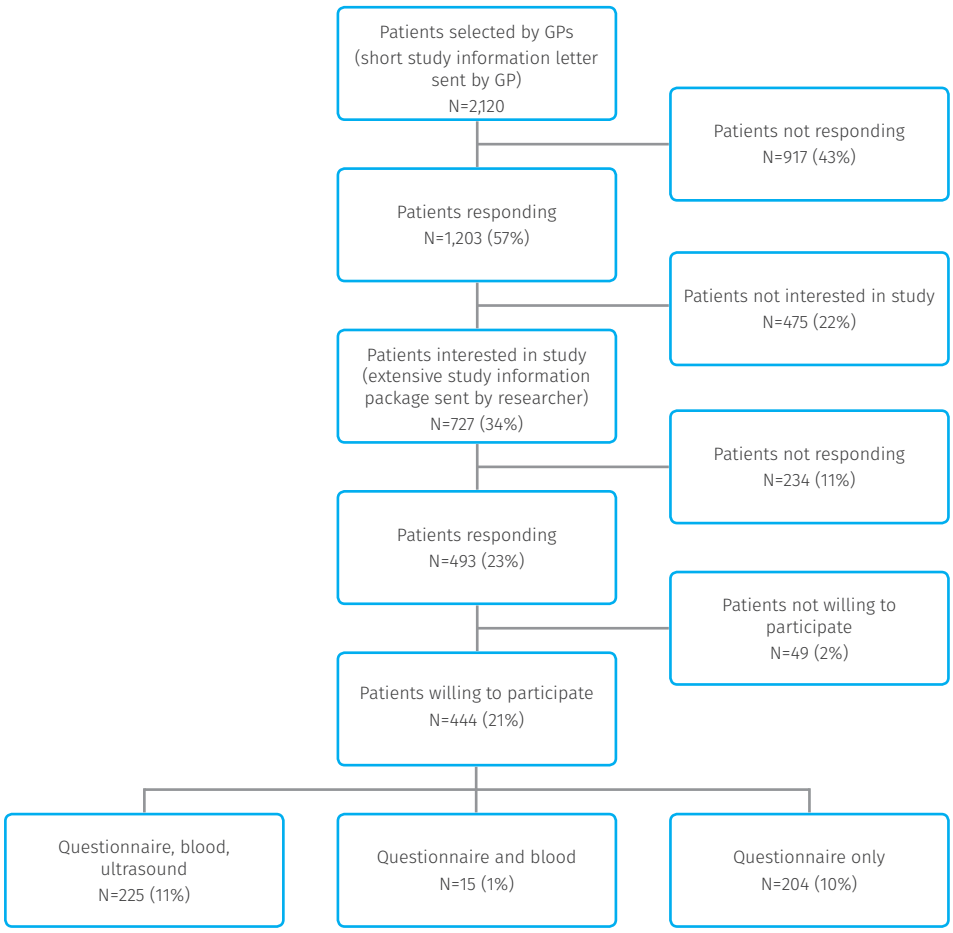


Figure 2 Response and participation rates of women recruited through the general practitioner (GP). Response and participation percentages are calculated using the initial number of selected patients (n=2,120) as the denominator.

sent by the general practitioner (the "non-responders"). From these latter two groups random samples of n=100 and n=200 women, were drawn, respectively. Subsequently, participants, refusers and non-responders were compared on the following characteristics: current age, parity, age at birth of first child, having experienced fertility problems in the past, and having consulted a gynecologist for these problems. For the participants group data from the study questionnaire were used, whereas for women in the refuser and non-responder groups, these data were retrieved from medical records available at the general practitioner's office.

No significant differences between the response groups were present regarding current, age, parity, age at birth of first child, having had fertility problems in the past, and having consulted a gynecologist in the past (Table 2), although the proportion of parous women was slightly lower among participants compared to refusers (46% vs. 58%; p = 0.06).

Table 2 Differences in fertility related characteristics between three types of response groups invited through the general practitioner

	Participants (group 1; n=194)	Refusers (group 2; n=100)	Non-responders (group 3; n=200)	P value grp 1 vs. grp 2	P value grp 1 vs. grp 3	P value grp 2 vs. grp 3
Current age; yrs (median (IQR))	34.0 (5.0)	35.0 (5.0)	34.0 (7.0)	NS	NS	NS
Parous, ever (N(%))*	89 (46)	54 (58)	96 (52)	NS	NS	NS
Age at birth first child (median (IQR))	32.0 (4.0)	32.0 (4.5)	32.0 (5.0)	NS	NS	NS
Has experienced fertility problems in the past (N(%))*	22 (12)	13 (14)	17 (9)	NS	NS	NS
Has consulted a gynecologist for fertility problems in the past (N(%))*	22 (12)	13 (14)	16 (9)	NS	NS	NS
Age at time of first gynecologist consultation (median (IQR))	31.0 (6.0)	35.0 (6.0)	34.0 (3.5)	NS	0.04	0.04

*IQR=interquartile range, NS = not statistically significant. * Total N does not correspond with the number mentioned in the heading of the table because of missing data.*

Within the group of women who had previously consulted a gynecologist for fertility problems, we assessed whether age at time of first consultation differed between the three response groups. It appeared that age at first consultation was significantly lower among participating women than among non-responding women (31.0 vs. 35.0 years, respectively; p = 0.04). It should be noted however, that these subgroups of women are rather small (i.e. participants n=22, non-responders n=16, refusers n=13). We concluded that, in general, selection bias due to non-response or non-participation was not present in our GP comparison group. The women invited through the general practitioner who participated in the DCOG LATER-VEVO study appear to represent the total group of invited women with respect to several important outcome characteristics.

Characteristics of CCSs, sisters, and GP controls

In Table 3 socio-demographic characteristics of the CCS group, the sister comparison group and the GP comparison group are presented.
Results show that participating CCSs were significantly younger than both the sister and GP controls (both p-values < 0.001). Moreover, the sister controls were younger than the GP controls (p < 0.001). With respect to educational level, we observed a significantly lower proportion of highly educated women in the CCS group compared to both the sister and the GP control groups (both p-values < 0.001). In addition, the proportion of highly educated women was significantly lower in sister controls compared to GP controls (p < 0.001). The marital status of CCSs was found to be

Table 3 Characteristics of CCSs, sister controls and GP controls participating in the DCOG LATER-VEVO study[#]

	CCSs (n=1051)	Sister Controls (n=357)	GP Controls (n=444)	P value CCSs vs. sisters	P value CCSs vs. GP controls	P value sisters vs. GP controls
Age; yrs (median IQR))	28.0 (12.1)	30.4 (11.8)	34.6 (11.8)	< 0.001	< 0.001	< 0.001
Educational level[†]						
Low	94 (9)	17 (5)	10 (2)	< 0.001	< 0.001	< 0.001
Medium	660 (63)	194 (55)	168 (38)			
High	290 (28)	142 (40)	264 (60)			
Marital status[†]						
Never married	737 (70)	213 (60)	281 (64)	0.001	0.01	NS
Married	291 (28)	134 (38)	148 (33)			
Divorced	20 (2)	8 (2)	11 (2)			
Widowed	0	0	2 (1)			
Employment status[†]						
Unemployed	161 (15)	25 (7)	30 (7)	< 0.001	< 0.001	NS
Student	187 (18)	43 (12)	32 (7)			
Employed	696 (67)	287 (81)	379(86)			

[#] Rates acquired as of January 1st 2014. Data presented as n/N (%) unless indicated otherwise. IQR=interquartile range.

[†] Total N does not correspond with the number mentioned in the heading of the table because of missing data. NS = not statistically significant.

'never married' at a significantly higher rate than both the sister and GP controls (p=0.001 and p=0.01, respectively). Whereas the sisters and GP controls did not differ significantly regarding marital status. Furthermore, CCSs appeared to be more often unemployed than sister and GP controls (both p-values < 0.001), while sisters and GP controls did not differ significantly with respect to their employment status. After correction for age, the significant differences between the three groups remained significant, except for the difference between CCSs and sister controls regarding marital status (see Supplementary Table 1).

It was evaluated whether the sister and GP controls differed regarding several basic fertility-related outcomes. Result showed that the rate of virginity did not differ significantly between both groups (25/346 (7%) vs. 20/440 (5%), respectively). In addition, no significant differences between sisters and GP controls were present regarding the proportion of women who had consulted a gynecologist for fertility problems in the past (33/335 (10%) vs. 46/423 (11%), respectively). The proportion of sisters and GP controls who have at least one child did not differ significantly either between both groups (141/347 (41%) vs. 207/443 (47%), respectively). After correction for age, all p-values remained non-significant.

DISCUSSION

In general, sibling controls are considered to be an attractive type of comparison group for late effect studies, as they share the same socio-demographic and genetic profile. However, the sample size of a sibling comparison group is limited, and this will become even more apparent as average household sizes are decreasing [165]. During our DCOG LATER-VEVO study it indeed became evident that the total number of sister controls would not be sufficient to address certain research questions of our study. Therefore, GP controls were also added to the comparison group of the study. However, the participation rate of GP controls was much lower than of sister controls (21% vs. 74%, respectively), whereas considerably more effort, and therefore higher costs, were involved in the recruitment of GP controls. The relatively high participation rate of sister controls compared to GP controls may not be surprising, since it is known that siblings of children with cancer are generally motivated to participate in research as they often share a special bond [166-168].

In our study, GP controls were recruited through general practitioners' offices, since we expected higher response rates than when using municipal registries as a sampling frame. Controls were recruited using a two-step method, in which interest for the study was evaluated in the first step, after which the extensive study information package was sent. This method appeared rather time-consuming and was associated with a high non-response. However, for practical as well as privacy reasons this method was the only available option for our DCOG LATER-VEVO-study. In fact, compared to the response and participation rates of a recent study on risk factors for Creutzfeldt-Jakob disease in the UK general population using the same method for recruiting control subjects, the response and participation rates in the current study can be considered fairly good [169].

For both the sister and the GP comparison group of the DCOG LATER-VEVO study, the risk of introducing selection bias was present. There was a clear personal benefit of participating in the study, as the participants received information regarding their reproductive function. This fact increased the risk of selection bias, as women with an active wish to become pregnant or women with fertility-related problems might be more likely to participate than women who already have children. In a pilot study conducted in 2005 (prior to the start of the DCOG LATER-VEVO study) we demonstrated that bias due to selective participation indeed was introduced. In this study CCSs were asked to invite their sister(s) for participation in the comparison group of the study but, in case no sisters were available or the survivor did not want her sister(s) to participate, female friends were allowed to participate as controls. The majority of the CCSs chose to ask female friends rather than their sisters to participate and the prevalence of fertility problems was higher among female friends of CCS than one would expect based on prevalence data in the general population [170]. Based on the results of this pilot study we decided for the DCOG LATER-VEVO study to recruit controls from the general population using a 'targeted' invitation strategy (i.e. by inviting women through the general practitioner only) rather than an open invitation strategy (i.e. using advertisements to recruit GP controls). It should be noted however, that all GP controls were recruited from the same geographic area. Moreover, almost half of the participating GP controls (194/444, 44%) were

recruited from one single GP office. People living in the same geographical often share comparable socio-demographic characteristics [171]. In addition, some of these characteristics, particularly education, are known to be related to impaired fertility [172]. Therefore, the fact that the geographical "coverage" of our GP controls was limited, could have led to a certain degree of bias in our study. However, no significant differences in reproductive characteristics were established between the women in the sister versus the GP comparison group with respect to rate of virginity, nulliparity and consultations with a gynecologist for fertility problems. Based on these results merging of both comparison groups for certain data analyses seems justified. In addition, and also of importance, is the fact that no significant differences in basic reproductive characteristics, such as parity, age at first child birth, and previous fertility problems and gynecologist consultations, were present between the participating, refusing, and non-responding women who were invited by the general practitioner. However, the GP controls were significantly older and higher educated than the sister controls (both P values < 0.001). These findings will be taken into account when analysing the data and interpreting the results of our study.

Altogether, it can be concluded that for the DCOG LATER-VEVO study sisters are considered a more appropriate type of comparison group than GP controls, as higher participation rates were acquired and the recruitment took considerably less effort. In addition, sisters are genetically similar to the CCSs, which is an additional advantage for some of our research questions. However, in our study the sisters appeared to differ from CCSs with respect to several socio-demographic variables, although in general these differences were smaller than the differences found between GP controls and CCSs. Another disadvantage encountered during our study was the fact that not all CCSs gave permission to contact their sisters, which may have led to a certain degree of bias and a limited number of sisters. Furthermore, we should bear in mind that sisters do not represent an unexposed group per se, as they might be psychosocially affected by the previous disease of their sister [164 173 174], which also may have impacted fertility-related outcomes. Finally, for some studies GP controls might be a more appropriate comparison group as certain outcomes can be compared with readily available reference norms, making it possible to easily check whether the comparison group accurately represents the general population [175]. Table 4 summarizes the several advantages and disadvantages of having siblings as well as subjects from the general population serving as controls in a study.

Table 4 Advantages and disadvantages of using different types of comparison groups in late effect studies

	Advantages	Disadvantages
Siblings	• High socio-demographic, environmental, and genetic comparability • Good participation rates	• Limited sample size • Siblings are not an unexposed group per se (i.e. they may also be affected by previous disease of sibling)
Subjects from general population	• Unlimited sample size • Readily available reference norms	• Limited participation rates • Risk of introducing selection bias • Recruitment can be expensive and time-consuming

For late effect studies among CCSs in general, one cannot designate one particular type of comparison group as the best to be used for these studies. Which type of comparison group to use depends on the research questions and outcomes of the study, the design of the study, and logistical, ethical and financial considerations. In any case, researchers setting up a study to investigate the occurrence of late effects among survivors of childhood cancer should carefully consider the advantages and disadvantages of choosing any type of comparison group.

APPENDIX

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.rbmo.2014.06.002

516278-L-bw-Overbeek
Processed on: 19-1-2018

5

Validity of self-reported data on pregnancies for childhood cancer survivors: a comparison with data from a nationwide population-based registry

A. Overbeek
M.H. van den Berg
C.W.P.M Hukkelhoven
L.C. Kremer
M.M. van den Heuvel-Eibrink
W.J.E. Tissing
J.J. Loonen
A.B. Versluys
D. Bresters
G.J.L. Kaspers
C.B. Lambalk
F. E. van Leeuwen
E. van Dulmen-den Broeder
on behalf of the DCOG
LATER-VEVO study group

Hum Reprod.
2013 Mar;28(3):819-27.

ABSTRACT

Study question
To what degree do records registered in the Netherlands Perinatal Registry (PRN) agree with self-report in a study questionnaire on pregnancy outcomes in childhood cancer survivors (CCSs)?

Summary answer
This study suggests that self-reported pregnancy outcomes of CCSs agree well with registry data and that outcomes reported by CCSs agree better with registry data than do those of controls.

What is known already
Many studies have shown that childhood cancer treatment may affect fertility outcomes in female CCSs; however, these conclusions were often based on questionnaire data, and it remains unclear whether self-report agrees well with more objective sources of information.

Study design, size, duration
In a nationwide cohort study on fertility (inclusion period January 2008 and April 2011, trial number: NTR2922), 1420 CCSs and 354 sibling controls were invited to complete a questionnaire regarding socio-demographic characteristics and reproductive history. In total, 879 CCSs (62%) and 287 controls (81%) returned the questionnaire.

Participants/materials, setting, methods
The current validation study compared the agreement between pregnancy outcomes as registered in the PRN and self-reported outcomes in the study questionnaire. A total of 589 pregnancies were reported in CCSs, and 300 pregnancies in sibling controls, of which 524 could be linked to the PRN.

Main results and the role of chance
A high intra-class correlation coefficient (ICC) was found for birthweight (BW) (0.94 and 0.87 for CCSs and controls, respectively). The self-reported BWs tended to be higher than reported in the PRN. For gestational age (GA), the ICC was high for CCSs (0.88), but moderate for controls (0.49). CCSs overestimated GA more often than controls. The Kappa values for method of conception and for method of delivery were moderate to good. Multilevel analyses on the mean difference with regard to BW and GA showed no differences associated with time since pregnancy or educational level.

Limitations, reasons for caution
Not all pregnancies reported could be linked to the registry data. In addition, the completeness of the PRN could not be assessed precisely, because there is no information on the number of missing records. Finally, for some outcomes there were high proportions of missing values in the PRN registry.

Wider implications of the findings
Our study suggests that questionnaires are a reliable method of data collection, and that for most variables, self-report agrees well with registry data.

INTRODUCTION

Advances in childhood cancer treatment over the past decades have significantly improved survival, resulting in a rapidly growing number of childhood cancer survivors (CCSs). However, in females there is evidence that both chemo- and radiotherapy may adversely affect reproductive function [39 176-178]. In studies conducted so far, primary outcomes considered in the field of female reproductive function of CCSs after anti-cancer treatment have included ovarian function, premature menopause, uterus function, actual fertility, pregnancy outcomes or a combination of these outcomes. The majority of studies have been large cohort studies in which the outcomes of interest were obtained through interviews or mailed questionnaires. However, the validity of such self-reported data may be limited, and may potentially lead to biased results.

Several studies have assessed the accuracy and validity of self-reported data regarding pregnancy outcomes in healthy women. It appears that correlations between medical charts and maternal self-report are high for birthweight (BW), gestational age (GA) and method of delivery [179 180]. Antenatal and perinatal complications and time to pregnancy, however, are reported less accurately [179-182].

We performed a validation study to assess the accuracy of pregnancy outcomes reported by CCSs and sibling controls in a mailed questionnaire.

METHODS

Study population

The current study is part of a Dutch nationwide study on reproductive function, ovarian reserve, premature menopause and pregnancy outcomes in CCSs, the so-called DCOG LATER-VEVO study. The study population, procedures and data collection methods have been described previously in detail [183]. In short, CCSs eligible for the nationwide study were selected from a cohort of patients treated for childhood cancer at one of the seven Dutch pediatric oncology and stem cell transplant centers between 1963 and 2002. The collaborative group for long-term effects after childhood cancer (LATER) has designed and implemented an electronic database, in each center, which includes patient and treatment details of all patients treated for cancer before the age of 18 years. The inclusion criteria for the DCOG LATER-VEVO study and the current study were identical and were defined as having been treated for a malignancy or central nervous system tumor before the age of 18, having survived for at least 5 years after diagnosis, being alive and being at least 18 years at study entry. The DCOG LATER-VEVO study was limited to female survivors. Women were excluded if they were not able to speak or read Dutch and if they had severe mental sequelae. All CCSs were asked for permission to invite their sister(s) for participation in the control group of the DCOG LATER-VEVO study. Eligible sisters were never diagnosed with cancer, had to be able to read and speak Dutch and had to be 18 years or older. The DCOG LATER-VEVO study consisted of three parts: a

questionnaire, the provision of a blood sample and a transvaginal ultrasound of the reproductive organs, with the two latter parts requiring a hospital visit. Eligible CCSs and sibling controls could decide either to refuse or to participate and, in case of participation, whether to take part in one, two or all three parts of the study. Data collection for the current study took place between January 2008 and April 2011. Women were eligible for the present study if they reported in the questionnaire ever having been pregnant. The exclusion criteria were pregnancy terminated before 24 weeks and date of birth of the offspring before 1 January 1985 or after 31 December 2009.

Data collection

For the purpose of the current study, only the questionnaire data were taken into account. The questionnaire is an adaptation of a well-tested questionnaire used by the Department of Epidemiology of the Netherlands Cancer Institute in a Dutch cohort study on long-term effects of ovarian stimulation for IVF [184 185]. The questionnaire addresses (amongst others) the following issues: socio-demographic characteristics, wish to have children and reproductive history, and for each pregnancy, data on maternal age, method of conception, duration of pregnancy, complications, pregnancy outcomes and details of the offspring.
Medical information regarding pregnancy outcomes was derived from the Netherlands Perinatal Registry (PRN), a nation-wide population-based registry, in which data of three medical registries (midwives, obstetricians and pediatricians/neonatologists) are combined. After delivery, standardized digital forms are entered in the nationwide database. These items are recorded by the caregiver, who is provided a standard manual with additional information on the definitions. The data are sent annually to the national registry office, where a number of range and consistency checks are conducted. False records are sent back to the caregiver with a request to correct them. Data in this registry are available from 1985 onwards. As from 1999, the PRN has included ~95% of all ~180 000 deliveries at > 16 completed weeks of gestation in the Netherlands [186]. The missing 5% is due to the fact that some midwives and general practitioners involved in obstetric care do not provide data for the PRN registry or because the records sent from the midwife practices are not received properly by the PRN. Before 1999, the PRN registry was less complete; however, exact proportions of completeness are not known as these records were not linked to Statistics Netherlands. In the PRN registry, the following variables are recorded: method of conception (natural, hormonal stimulation, IUI, controlled ovarian hyperstimulation, IVF and other), method of delivery, BW (in grams), GA (in weeks) and highest diastolic blood pressure (in mmHg). In the DCOG LATER-VEVO questionnaire, women reported on method of conception (natural, hormonal stimulation, IUI, IVF/ICSI), complications during pregnancy (hypertension and growth retardation of the child), method of delivery (spontaneous, assisted delivery, induction of labor, Caesarean section), date of birth of the baby, GA, sex and BW. All participants gave written informed consent for data abstraction from medical records. The records of self-reported pregnancies from the questionnaire

were linked to the PRN by using both the mother's date of birth and the child's date of birth as linkage keys. If linkage led to multiple hits, the sex of the baby and the postal code of the mother were used in an attempt to correctly link self-reported data with PRN data. If it was not possible to link self-reported data with a unique corresponding record in the PRN, the data were not included in the current study.

Statistical analysis

The data were checked for a normal distribution. Data of continuous variables are presented as the mean (standard deviation (SD)) if normally distributed or as the median and interquartile range (IQR) if not normally distributed. In case CCSs and controls were compared, an independent Student's t-test was used when data were normally distributed. Mann–Whitney U-test was used to compare CCSs and controls when data were not normally distributed. We calculated intra-class correlation coefficients (ICCs) and confidence intervals for the continuous variables, BW and GA. ICCs were calculated using a two-way random effects model. BW and GA were categorized (BW: very low birthweight (< 1500 g), low birthweight (1500–2500 g), normal birthweight (2500–4000 g) and high birthweight (> 4000 g); GA: preterm (< 37 weeks), term (37–42 weeks), and post-term (> 42 weeks)). For BW, GA and method of delivery, the PRN was considered the gold standard. Sensitivity, as well as specificity, was calculated. Sensitivity was defined as the proportion of those with the condition (as defined by the PRN) who are correctly classified by the questionnaire. Specificity was the proportion of those without the condition (as defined by the PRN), who are correctly classified by the questionnaire. For method of conception and pregnancy complications, the PRN could not serve as the gold standard, since these variables are not registered consistently. Therefore, for these variables, only reliability measures were calculated. In order to assess the agreement between self-reported data and PRN data for categorical variables, the proportion of overall agreement and Cohen Kappa statistics were calculated. The proportion of overall agreement, which is the proportion of cases for which PRN and self-report agree, is a crude descriptive measure that is informative, useful and easy to interpret, but does not distinguish between agreement on positive ratings and agreement on negative ratings and does not take into account chance. Therefore we also calculated Kappa values. According to Landis and Koch, variables with values of Kappa > 0.75 can be considered as an excellent agreement, values of 0.40–0.75 as a moderate agreement, and values below 0.40 as a poor agreement [187]. In case a woman had reported more than one pregnancy, only the first reported pregnancy was used for the calculation of validity and reliability outcomes in order to avoid dependency of observations. To determine which variables were independently associated with overall agreement, multilevel analysis was performed, allowing the correction for the clustering of pregnancies for one woman. Variables included in this analysis were time between delivery and questionnaire, maternal age and educational level. Analyses were performed using SPSS software (version 15.0, SPSS, Inc., Chicago, IL) and MLWin (version 2.24, Centre for Multilevel Modelling, University of Bristol).

RESULTS

During the inclusion period of the current study, 1,420 CCSs and 354 sibling controls were invited for the DCOG LATER-VEVO study. In total, 879 CCSs (62%) and 287 controls (81%) returned the DCOG LATER-VEVO questionnaire. In 289 CCSs, 589 pregnancies were reported and 300 pregnancies were reported in 123 controls. There were 160 pregnancies in the survivor group and 67 pregnancies in the control group that had to be excluded because the date of birth was before 1 January 1985 or after 31 December 2009 or unknown and/or the pregnancy was terminated before 24 weeks of GA. This resulted in 429 pregnancies of CCSs and 233 pregnancies of controls included in the study. Linkage to the PRN database yielded 488 unique hits. In 37 cases linkage led to multiple hits in the PRN. All but one could subsequently be identified by the sex of the baby and/or by the postal code of the mother. No matching records were found in the PRN database for 72 pregnancies in the survivor group and 65 pregnancies in the control group. Finally, 357 pregnancies reported by 218 CCSs and 167 pregnancies reported by 105 controls were included in the data-analyses (Figure 1).

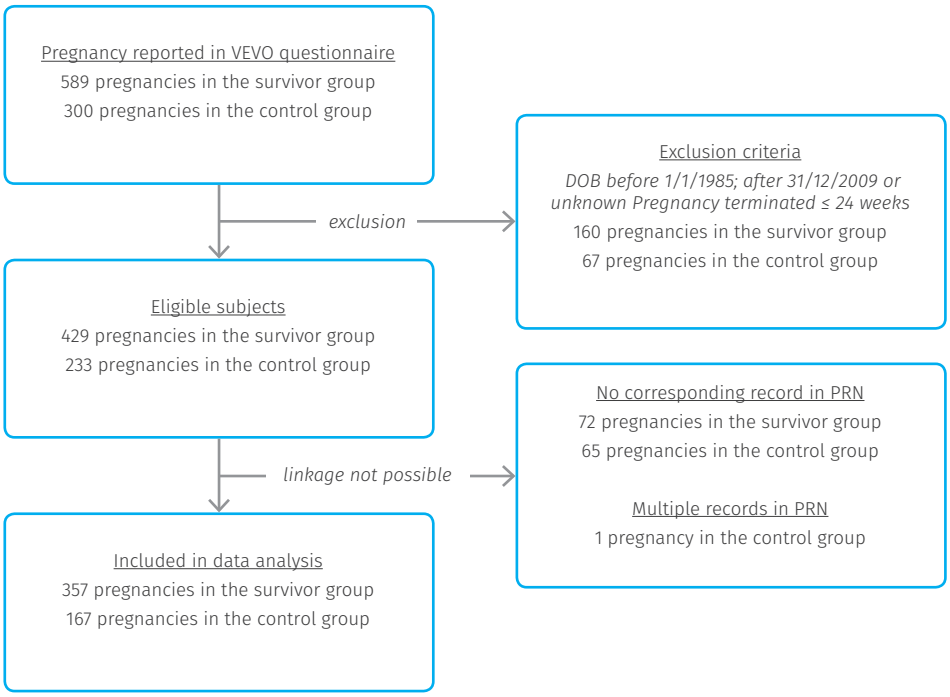


Figure 1 Description of the recruitment of eligible women from the DCOG LATER-VEVO study population. DOB; date of birth.

There were 78 deliveries (28%) of CCSs and 49 pregnancies (29%) of controls that took place before 1999. Of those that could not be linked, 22 (35%) and 17 (39%) pregnancies of CCSs and controls, respectively, took place before 1999. Table 1 presents the basic characteristics of the CCSs and the controls for the included

Table 1 Characteristics of participants of the DCOG LATER-VEVO study included in the validation study, the Netherlands, 2008-2011

Characteristic	Included in validation study				Record not linked with PRN			
	CCS (n = 218)		Controls (n = 105)		CCS (n = 63)		Controls (n = 44)	
	N	%	N	%	N	%	N	%
Age (years) at questionnaire								
< 30	58	26.6	19	18.1	15	23.8	3	6.8
30-35	66	30.3	29	27.6	13	20.6	14	31.8
35-40	54	24.8	30	28.6	18	28.6	16	36.4
> 40	40	18.3	27	25.7	17	27.0	11	25.0
Maternal age (years) at first born								
< 25	44	20.2	25	23.8	14	22.2	11	25.0
25-30	107	49.1	39	37.1	29	46.0	16	36.4
30-35	59	27.1	3	31.4	16	25.3	15	34.1
> 35	8	3.7	8	7.6	4	6.3	2	4.5
Duration (years) from child birth to questionnaire								
< 2	50	22.9	20	19	17	27.0	5	11.4
2-5	69	31.7	22	21	12	19.0	11	25
5-8	38	17.4	23	21.9	8	12.7	9	20.5
> 8	61	28.0	39	37.1	26	41.3	19	43.1
Maternal educational level*								
Lowest	5	2.3	0	0	0	0	0	0
Low	38	17.4	15	14.3	7	11.1	5	11.4
Medium	103	47.2	46	43.8	31	49.2	19	43.1
High	72	33.0	44	41.9	25	39.7	20	45.5

**Educational level: lowest (special education), low (primary school), medium (secondary school), high (college or university).*

group as well as of those who could not be linked to the PRN records.
The mean ages at completion of the questionnaire (SD) were 34.5 (6.3) and 36.7 (7.0) years for CCSs and controls, respectively, for those who could be included in the validation study (P = 0.03) and 36.8 (7.0) and 38.1 (6.5) years for CCSs and controls who could not be linked to PRN records, respectively. The mean maternal age at delivery (SD) was 29.2 (4.3) years for CCSs, whereas controls were 28.8 (4.0) years (P = 0.47), comparable to the age of those who could not be linked, namely 28.4 (5.1) and 28.3 (4.1) for CCSs and controls, respectively. The median durations (IQR) from child birth to questionnaire were 4.7 (10.7) and 5.5 (14.5) for CCSs and controls, respectively (P = 0.05), whereas in the non-linked group these were 7 (10) and 7 (9.8) for CCSs and controls, respectively.

The ICC between self-report and registry regarding BW was high (0.94 (95% confidence interval: 0.91-0.96) and 0.87 (95% confidence interval: 0.83-0.90) for CCSs and controls, respectively). For GA, the ICC was also high for CCSs (0.88, 95% confidence interval: 0.85-0.91), but moderate for controls (0.49, 95% confidence interval: 0.32-0.62).
In Tables 2 and 3, the results of the comparison between self-reported data and

Table 2 Comparison of the self-reported data from the DCOG LATER-VEVO questionnaire with data from the Netherlands Perinatal Registry regarding various pregnancy outcomes in childhood cancer survivors

Birth Weight	**Reported in Netherlands Perinatal Registry**					**Agreement**	**Kappa value**	**95% CI**
Reported in questionnaire	Very low BW (< 1500g)	Low BW (1500g -2500g)	Normal BW (2500g -4000g)	High BW (> 4000g)	Missing			
Very low BW (< 1500g)	3	0	0	0	0	98	0.59	0.23-0.95
Low BW (1500g-2500g)	0	19	4	0	0	98	0.87	0.76-0.98
Normal BW (2500g-4000g)	0	1	145	0	0	91	0.79	0.70-0.88
High BW (> 4000g)	4	0	14	28	0	92	0.71	0.59-0.83

BW = birth weight

Gestational age	**Reported in Netherlands Perinatal Registry**				**Agreement**	**Kappa value**	**95% CI**
Reported in questionnaire	Preterm (< 37 weeks)	Term (37-42 weeks)	Post term (> 42 weeks)	Missing			
Preterm (< 37 weeks)	40	6	0	0	93	0.79	0.69-0.89
Term (37-42 weeks)	10	159	0	0	91	0.75	0.65-0.86
Post term (> 42 weeks)	0	3	0	0	99	n/a	n/a

Method of conception	**Reported in Netherlands Perinatal Registry**					**Agreement**	**Kappa value**	**95% CI**
Reported in questionnaire	Natural	Hormonal	IUI	IVF/ICSI	Missing			
Natural	151	2	0	0	43	97	0.82	0.67-0.97
Hormonal	1	2	0	0	2	98	0.56	0.12-1.01
IUI	0	0	3	1	3	99	0.85	0.57-1.14
IVF/ICSI	2	0	0	7	1	98	0.81	0.61-1.02

Method of delivery	Reported in Netherlands Perinatal Registry				Agreement	Kappa value	95% CI
Reported in questionnaire	Spontaneous	Vacuum/ forcipal extraction	Caesarean section	Missing			
Spontaneous	96	8	1	21	88	0.76	0.67-0.85
Vacuum/ forcipal extraction	2	41	0	1	95	0.86	0.77-0.94
Caesarean section	12	0	36	0	93	0.81	0.71-0.91

Pregnancy complications	Reported in Netherlands Perinatal Registry			Agreement	Kappa value	95% CI
Reported in questionnaire	Hypertension	No hypertension	Missing			
Hypertension	28	17	3	90	0.59	0.45-0.73
No hypertension	10	136	24	90	0.59	0.45-0.73

PRN data regarding various categorical pregnancy outcomes are presented. For the categories of low and normal BW, the Kappa values in the survivor group were high (0.87 and 0.79, respectively); however, the corresponding values were moderate in the control group (0.42 and 0.61, respectively). The high BW category was scored

Table 3 Comparison of the self-reported data from the DCOG LATER-VEVO questionnaire with data from the Netherlands Perinatal Registry regarding various pregnancy outcomes in controls

Birth Weight	Reported in Netherlands Perinatal Registry					Agreement	Kappa value	95% CI
Reported in questionnaire	Very low BW (< 1500g)	Low BW (1500g -2500g)	Normal BW (2500g -4000g)	High BW (> 4000g)	Missing			
Very low BW (<1 500g)	1	0	0	0	0	100	1.00	1.00-1.00
Low BW (1500g-2500g)	0	2	3	0	0	95	0.42	0.004-0.84
Normal BW (2500g-4000g)	0	1	73	0	0	86	0.61	0.44-0.78
High BW (> 4000g)	0	1	11	13	0	89	0.62	0.44-0.81

BW = birth weight

Gestational age	Reported in Netherlands Perinatal Registry				Agreement	Kappa value	95% CI
Reported in questionnaire	Preterm (< 37 weeks)	Term (37-42 weeks)	Post term (> 42 weeks)	Missing			
Preterm (< 37 weeks)	9	4	0	0	94	0.72	0.50-0.93
Term (37-42 weeks)	2	87	0	1	92	0.65	0.43-0.87
Post term (> 42 weeks)	0	2	0	0	98	n/a	n/a

Method of conception	Reported in Netherlands Perinatal Registry					Agreement	Kappa value	95% CI
Reported in questionnaire	Natural	Hormonal	IUI	IVF/ICSI	Missing			
Natural	65	1	0	0	32	96	0.71	0.39-1.02
Hormonal	0	1	0	0	1	99	0.66	0.04-1.28
IUI	1	0	0	0	0	99	n/a	n/a
IVF/ICSI	1	0	0	3	0	99	0.85	0.56-1.14

Method of delivery	Reported in Netherlands Perinatal Registry				Agreement	Kappa value	95% CI
Reported in questionnaire	Spontaneous	Vacuum/ forcipal extraction	Caesarean section	Missing			
Spontaneous	50	0	0	25	94	0.86	0.75-0.98
Vacuum/ forcipal extraction	1	14	0	0	98	0.92	0.81-1.03
Caesarean section	4	1	10	0	94	0.77	0.57-0.96

Pregnancy complications	Reported in Netherlands Perinatal Registry			Agreement	Kappa value	95% CI
Reported in questionnaire	Hypertension	No hypertension	Missing			
Hypertension	8	6	1	90	0.61	0.37-0.85
No hypertension	2	64	24	90	0.61	0.37-0.85

with a moderate agreement (0.71 and 0.62 in CCSs and controls, respectively). BWs reported in the questionnaire tended to be higher than those reported in the PRN. Of the 23 discrepancies in the survivor group, BW was overestimated by self-report in 19 cases (83%). In the control group, self-report overestimated the BW category in 13 of the 16 cases (81%). For this category, there were no missing values. In both groups, categories of GA were reported with a moderate agreement (preterm delivery: 0.79 and 0.72; term delivery: 0.75 and 0.65, for CCSs and controls, respectively). When evaluating discrepancies between self-report and registry data, GA in the CCSs group was often overestimated in the self-reported questionnaire (13 of 19 cases (68%)). However, within the control group GA was overestimated as frequently as underestimated (4 versus 4 cases).

The Kappa value for method of conception varied largely per method (0.82 and 0.71 for natural conception, 0.56 and 0.66 for controlled ovarian hyperstimulation, 0.81 and 0.85 for IVF/ICSI, for CCSs and controls, respectively). When examining the discrepancies between self-report and PRN with regard to method of conception, no specific direction of misclassification could be discerned in the survivor group or in the control group. Moreover, 49 of the 218 cases (22%) in the survivor group were missing in the PRN registry, 43 of which were self-reported as natural conceptions. For this category in the control group, 33 of 105 cases (31%) were missing in the PRN, of which 32 cases were self-reported natural conception.

For the method of delivery, the Kappa values were 0.76 and 0.86 for spontaneous delivery, 0.86 and 0.92 for vacuum or forcipal extraction and 0.81 and 0.77 for Caesarean section in CCSs and controls, respectively. CCSs reported the method of delivery as spontaneous in nine cases, whereas PRN stated additional techniques were used to enable the delivery (vacuum or forcipal extraction or even Caesarean section). Conversely, in 14 cases, the survivor reported an assisted delivery, while PRN records showed a spontaneous delivery. Of the six discrepancies between PRN and self-report regarding the method of delivery in the control group, all were due to controls reporting an assisted delivery, while PRN mentioned a spontaneous delivery. With regard to the method of delivery, 22 cases in the survivor group (10%) and 25 cases in the control group (24%) were missing in the PRN data, 21 (95%) and 25 (100%) of which, respectively, were self-reported spontaneous deliveries. The Kappa values for pregnancy-induced hypertension were 0.59 for CCSs and 0.61 for controls. For this variable there were 27 discrepancies between PRN and self-report in the survivor group and 8 in the control group, the direction of which was non-specific. In the survivor group, 27 records were missing in the PRN, 3 of which were reported as hypertension by self-report. In the control group, 25 records were missing in the PRN registry, one of which was reported as hypertension in the questionnaire.

In Table 4, validity measures are presented for BW, GA and method of delivery. Sensitivity ranged from 42.9 to 100% in CCSs, and from 33.3 to 100% in controls. Specificity was good, ranging from 89.5 to 100% in CCSs and from 81.8 to 100% in controls. In addition, the influence of educational level, maternal age and the time between delivery and questionnaire on the difference in BW and GA between self-

reported and registered data was assessed by multilevel analysis. None of these factors were associated with a less accurate recall of pregnancy outcomes. More specifically, there were no large differences between self-reported and registered data regarding BW and GA with longer time since birth, lower educational level or higher maternal age.

Table 4 Measures of validity of birth weight, gestational age and method of delivery

	CCS			Controls		
	PPV (%)	Sensitivity (%)	Specificity (%)	PPV (%)	Sensitivity (%)	Specificity (%)
Birth weight						
Very low (< 1500g)	100	42.9	100	100	100	100
Low (1500g-2500g)	82.6	95.0	98.0	40.0	33.3	97.0
Normal (2500g-4000g)	99.3	89.0	98.2	98.7	83.9	94.4
High (> 4000g)	60.9	100	90.5	52.0	100	87.0
Gestational age						
Preterm (< 37 weeks)	87.0	80	96.4	69.2	81.8	95.7
Term (37-42 weeks)	94.1	94.6	80.0	97.8	93.6	81.8
Post term (> 42 weeks)	n/a	n/a	98.0	n/a	n/a	98.1
Method of delivery						
Spontaneous	91.4	87.2	89.5	100	90.9	100
Vacuum/forcipal extraction	95.4	83.7	98.6	93.3	93.3	98.5
Caesarean section	75.0	97.3	92.5	66.7	100	92.7

Data from the Netherlands Perinatal Registry were considered gold standard. PPV; positive predictive value.

DISCUSSION

In this study, the agreement between self-reported pregnancy outcomes and registry data in female CCSs and controls was examined. Our results show that the validity of pregnancy outcomes reported by CCSs is good for GA and BW. However, sibling controls reported GA with only a moderate agreement. CCSs as well as controls tended to overestimate BW and CCSs more often overestimated GA. Hypertension in pregnancy and controlled ovarian hyperstimulation were reported with moderate-to-good agreement, whereas the method of conception and delivery was reported with a good agreement. The sensitivity and specificity were high for CCSs and controls for BW, GA and method of delivery. A higher maternal age at childbirth, a longer time since pregnancy and a lower educational level were not associated with lower agreement in either group. Overall, it seemed that some self-reported pregnancy outcomes of CCSs (specifically birth weight and gestational age) agreed better with the registry data than those reported by controls, indicating a potential source of

differential misclassification. This might be due to the increased awareness of late effects and a higher frequency of medical follow-up among CCSs.
To our knowledge, no studies have been conducted to assess the validity of self-reported data on pregnancy outcomes in CCSs. However, with respect to the self-report of late effects in general, it has been found that CCSs report a wide range of late effects in significantly greater numbers than recorded in medical notes [188 189]. Furthermore, CCSs appear to show a biased response style, indicating a systematic tendency to deny difficulties on QOL measures [190]. Olson et al. conducted a study in which mothers of CCSs were interviewed on pregnancy and delivery information. The authors concluded that the validity and reliability of maternally reported pregnancy and delivery information may vary with the nature of the factor of interest (i.e. BW, pregnancy complications etc.), but is affected little by time from birth or case–control status [179]. In concordance with our study, high correlations for BW and GA were found. However, self-report by mothers of CCSs on pregnancy-induced hypertension and method of delivery had low validity and reliability scores, whereas in our study, there seem to be moderate-to-high agreement scores. The study described by Olson et al. took place in the 80s, whereas our study was conducted > 20 years later. The differences in the validity of self-reported data on pregnancy complications and medical interventions may be due to the different time frame in which both studies were conducted. Possibly, the level of communication between the patient and her physician has changed in this period, leading to a better recall of pregnancy complications and medical interventions in our study. Finally, the way in which the method of delivery and medical conditions during pregnancy were questioned may have differed between the study of Olson et al. and our study, which may also have caused differences in accuracy. Rice et al. [180] evaluated the agreement between maternal report and medical records in women who gave birth to a child following IVF. Correlations between self-report and medical records for BW were comparable to our results. The Kappa value for delivery via Caesarean section, however, was 1.00, in contrast to 0.81 in our study. In the study of Rice et al., a distinction was made between Caesarean sections in general and emergency Caesarean sections. The Kappa value for the emergency Caesarean sections in the study of Rice et al. was 0.78. Moreover, the authors pointed out that their results may overestimate recall rates of pregnancy outcomes in the general population, because women who were pregnant after IVF treatment may better recall pregnancy-related outcomes [180]. In accordance with the study of Rice et al., Tomeo et al. and Sou et al. described a validity of 100% for the report of Caesarean sections in a group of women recruited from the general population, whereas induction of labor and forcipal or vacuum extraction were reported less sensitively (sensitivity and specificity of 93 and 86% for induction of labor and 26 and 74% for assisted delivery) [191 192].
When interpreting our results, the limitations of this study should be considered. First, we were not able to link 17% (72/429) of the pregnancies in the survivor group and 28% (65/233) in the control group. In other recent studies, higher linkage rates have been reached [CWPM Hukkelhoven, Perinatal Registry Utrecht, personal communication, 2012]. In these studies, in which recent pregnancies are evaluated, not only date of birth of the child, but also BW, GA and postal code could be used as linkage keys. Due to the objective of our study, i.e. investigating whether self-

reported pregnancy outcomes such as BW and GA agree with those in the registry, it was not possible to use these outcomes as linkage keys. Postal code could often not be used, as it was likely that participants who reported a pregnancy from many years ago may have changed address since they were pregnant. It could also be that due to input errors of birth dates, either in the registry or in the self-reported data, some pregnancies could not be linked. Secondly, the completeness of the PRN could not be assessed precisely, because there is no information on the number of missing records. Records in the PRN database agree with records of Statistics Netherlands for 95% and slightly less in earlier years [186]. However, the database of Statistics Netherlands can also not be considered complete, as it contains no data on births among women who stay illegally in the Netherlands. In this study, pregnancy records were available from 1985 onwards, and it may very well be possible that many of the older records in the PRN could not be linked due to incompleteness of the registry or due to recall bias of the participants. Indeed, the proportion of records in which date of birth was reported before 1999 was larger for non-linked records than for linked records. Thirdly, for BW, GA and method of delivery, the PRN database was considered the gold standard. However, 12% of the records of the CCSs and 24% of those of the controls were missing in the PRN with regard to the method of delivery. Agreement on Caesarean section was only 0.77 in controls and 0.81 in CCSs. In other studies, recall of Caesarean section often agrees for the full 100% [180 191 192]. It is unlikely that women forget (to report) such an operation. As Kappa values are < 0.8, one could therefore cautiously question the quality of the PRN registry. Unfortunately, input errors in the PRN cannot be excluded, nor can they be quantified, as, for this study, there was no alternative source of pregnancy records available. Finally, for many cases, the categories of the outcomes 'method of delivery' and 'method of conception' had missing values in the PRN. This can be explained by the fact that providing information on method of conception was only mandatory for gynecologists; for midwives this was optional. This may lead to a differential misclassification, as pregnancies under gynecologic supervision are often more complicated than those supervised by midwives. In this way, it may seem that an alternative method of conception, other than spontaneous, influences pregnancy outcomes and complications.

In conclusion, our results indicate that for the most important outcomes regarding fertility and pregnancy, self-report in CCSs is consistent with registry parameters. The use of questionnaires in CCSs to assess pregnancy outcomes therefore seems justified. However, since we observed differences in accuracy between CCSs and controls, differential misclassification should be considered when interpreting the data. In addition, one should realize that especially BW and GA are more often overestimated in self-reported questionnaires.

516278-L-bw-Overbeek
Processed on: 19-1-2018

6

Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics

M.H. van den Berg
A. Overbeek
H.J. van der Pal
A.B. Versluys
D. Bresters
F.E. van Leeuwen
C.B. Lambalk
G.J.L. Kaspers
E. van Dulmen-den Broeder

J Med Internet Res.
2011 Sep 29;13(3):e76.

ABSTRACT

Background
Web-based questionnaires have become increasingly popular in health research. However, reported response rates vary and response bias may be introduced.

Objective
The aim of this study was to evaluate whether sending a mixed invitation (paper-based together with Web-based questionnaire) rather than a Web-only invitation (Web-based questionnaire only) results in higher response and participation rates for female childhood cancer survivors filling out a questionnaire on fertility issues. In addition, differences in type of response and characteristics of the responders and nonresponders were investigated. Moreover, factors influencing preferences for either the Web- or paper-based version of the questionnaire were examined.

Methods
This study is part of a nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female childhood cancer survivors. The Web-based version of the questionnaire was available for participants through the Internet by means of a personalized user name and password. Participants were randomly selected to receive either a mixed invitation (paper-based questionnaire together with log-in details for Web-based questionnaire, n = 137) or a Web-only invitation (log-in details only, n = 140). Furthermore, the latter group could request a paper-based version of the questionnaire by filling out a form.

Results
Overall response rates were comparable in both randomization groups (83% mixed invitation group vs 89% in Web-only invitation group, P = 0.20). In addition, participation rates appeared not to differ (66% or 90/137, mixed invitation group vs 59% or 83/140, Web-only invitation group, P = 0.27). However, in the mixed invitation group, significantly more respondents filled out the paper-based questionnaire compared with the Web-only invitation group (83% or 75/90 and 65% or 54/83, respectively, P = 0.01). The 44 women who filled out the Web-based version of the questionnaire had a higher educational level than the 129 women who filled out the paper-based version (P = 0.01). Furthermore, the probability of filling out the Web-based questionnaire appeared to be greater for women who were allocated to the Web-only invitation group (OR = 2.85, 95% CI 1.31 - 6.21), were older (OR = 1.08, 95% CI 1.02 - 1.15), had a higher educational level (OR high vs low = 0.06, 95% CI 0.01 - 0.52), or were students (OR employed vs student = 3.25, 95% CI 1.00 - 10.56).

Conclusions
Although overall response as well as participation rates to both types of invitations were similar, adding a paper version of a questionnaire to a Web-only invitation resulted in more respondents filling out the paper-based version. In addition, women who were older, had a higher level of education, or were students, were more likely to have filled out the Web-based version of the questionnaire. Given the many advantages of Web-based over paper-based questionnaires, researchers should strongly consider using Web-based questionnaires, although possible response bias when using these types of questionnaires should be taken into account.

INTRODUCTION

The number of Internet users worldwide has doubled in the past 5 years, and it is estimated there are over 2 billion users in 2010 [193]. In the Netherlands, there were up to 15 million Internet users in 2010, representing 85.6% of the Dutch population [194]. Not surprisingly, within research settings, the Internet is increasingly being used as a tool for collecting data by means of Web-based questionnaires. The use of this type of questionnaire is less time-consuming and less costly compared with the use of paper-based questionnaires. For Web-based questionnaires, no printing and mailing costs are involved and the time spent by a researcher on data entry is minimal since the returned data are already in an electronic format. In addition, studies have reported that Web-based questionnaires have fewer response errors and fewer socially desirable responses, while no differences have been found in the accuracy of the reported information between the two types of questionnaires [195-197]. Thus, Web-based questionnaires might serve as an attractive alternative to paper-based questionnaires, especially when the target study population primarily consists of relatively young respondents [198]. However, important technical and methodological issues have been raised that should be carefully considered when using Web-based questionnaires [199 200]. An important issue is obtaining representative samples of the study population with adequate response rates to secure external validity. If response rates to a Web-based questionnaire appear to be low or seem to come from a selective group, response bias is introduced and the results might be misleading [198 201].

In the past decade, several studies have investigated response rates of Web-based versus paper-based questionnaires in many different populations and in many different settings. Response rates appear to vary widely and seem to be more dependent on the population sampled than on any other factor [202-204]. Although conflicting results have been published, recent studies have demonstrated an increase in response rates to web-based compared with paper-based questionnaires [203-206]. Employing a mixed-mode strategy, enabling patients to fill out either a Web-based or a paper-based questionnaire, seems to enhance response rates even further [207 208]. These two types of questionnaires can be offered simultaneously or sequentially, a factor that also seems to influence response rates [209].

The current study is part of a Dutch nationwide study. This study, which was initiated in 2008, examines the effects of childhood cancer and its treatment on reproductive function, ovarian reserve, and risk of premature menopause in female childhood cancer survivors. The questionnaire used in this study, of which both a paper-based version as well as a web-based version was available, contains questions about several fertility-related issues. So far, no studies are available comparing response rates to a Web- and paper-based version of a questionnaire on fertility issues among young adult female cancer survivors. Indeed, previous studies among female childhood cancer survivors have predominantly used paper-based questionnaires, telephone interviews, and face-to-face interviews to collect data [210]. In the US Childhood Cancer Survivor Study (CCSS) as well as in the UK British Childhood Cancer Survivor Study (BCCSS), paper-based questionnaires have been sent to large cohorts of childhood cancer survivors. These questionnaires contained questions about

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 107

sociodemographic items, adverse health outcomes, use of medications, lifestyle behavior, pregnancy history, and family history. Reported response rates were 82% (CCSS) and 71% (BCCSS), respectively [211 212].

In addition, studies investigating response rates to Web-based questionnaires among survivors of childhood cancer are scarce. Thompson et al. [213] used a Web-based questionnaire to investigate difficulties regarding romantic relationships in childhood cancer survivors. For this purpose, 603 survivors were sent a letter by postal mail inviting them to participate. Only 60 survivors (10%) agreed to participate and filled out the Web-based questionnaire. Low response rates were also reported by Cantrell et al. [214] in their study of health-related quality of life following childhood cancer. A Web-based survey was used, which was brought to the attention of potential respondents by posting a link on six different websites intended for use by survivors of childhood cancer. Although exact response rates could not be calculated, the authors reported the response rate to be low and the time needed to collect data to be long. In another study, childhood cancer survivors were recruited for a Web-based survey on physical activity through advertisements posted on cancer survivor–related websites and newsletters [215]. Since that study also used a reactive recruitment method, no true response rates could be calculated. However, the authors stated that they realize that the generalizability of their study was limited, as the recruitment method used probably had led to a specific group of survivors responding to the study invitation.

In conclusion, it is not known what response and participation rates can be expected when inviting female childhood cancer survivors to fill out a Web-based or a paper-based questionnaire on fertility issues. More specifically, no information is available on the impact of adding a paper-based questionnaire to an invitation to fill out the Web-based questionnaire. Therefore, we aimed to evaluate whether sending a mixed invitation (paper-based together with Web-based questionnaire) rather than a Web-only invitation (Web-based questionnaire only) results in higher response and participation rates for female childhood cancer survivors filling out a questionnaire on fertility issues. Furthermore, in order to identify possible response bias, differences in type of response and characteristics of the responders and nonresponders were investigated. Moreover, factors influencing preferences for either the Web- or paper-based version of the questionnaire were examined.

METHODS

Eligible survivors for the nationwide study were selected from a cohort of patients treated for childhood cancer at one of the five Dutch pediatric oncology centers or one of the two stem cell transplant centers between 1965 and 2002. Within the collaborative Dutch Childhood Oncology Late Effects Group, an electronic database has been set up in each center that includes patient and treatment details of all patients treated for cancer before the age of 18 years. Inclusion criteria for the nationwide study and the current study were identical and were defined as: female sex, having been treated for a malignancy or central nervous system tumor before

the age of 18, having survived for at least 5 years, being alive, and being at least 18 years of age at study entry. Patients were excluded if they were not able to speak or read Dutch or if they had severe sequelae related to mental health.
The nationwide study consists of three components: a questionnaire, the provision of a blood sample, and a transvaginal ultrasound measurement of the reproductive organs. The last two of these components require a hospital visit. Patients can either refuse to participate or take part in one, two, or all three components of the study. For the purpose of the current report, only the questionnaire component was taken into account.

Questionnaire and procedures for distribution

The questionnaire used in the study is an adaptation of a well-tested questionnaire used by the Department of Epidemiology of the Netherlands Cancer Institute in a large-scale Dutch cohort study on long-term effects of ovarian stimulation for in vitro fertilization [216]. It addresses the following issues: sociodemographic characteristics, medical history, menstrual and reproductive history, pregnancy outcomes, menopausal symptoms and menopause, and family history of cancer and family history of subfertility or infertility.
The paper- and Web-based version of the questionnaire were identical in terms of the questions asked, their wording, and their order of presentation. In the Web-based version, however, questions not relevant to the participant were automatically skipped. The Web-based version of the questionnaire was accessible for participants through a website which was specially designed for the nationwide study.
The study population for this study consisted of 277 female childhood cancer survivors from three participating centers of the nationwide study (Emma Children's Hospital/ Academic Medical Center Amsterdam, Leiden University Medical Center, University Medical Center Utrecht/Wilhelmina Children's Hospital). These women were randomly allocated to two groups: the mixed invitation group and the Web-only invitations group.

The mixed invitation group

Participants in the mixed invitation group received an invitation that contained a paper-based questionnaire together with an instruction sheet for the Web-based questionnaire. This instruction sheet contained a personalized username, the name of the website, and a log-in code allowing them to log in to a secured part of the website and fill out the questionnaire.

The web-only invitation group

Participants in the Web-only group received the above-mentioned instruction sheet containing the name of the website and the log-in details alone. However, a paper-based questionnaire could be acquired by ticking this option on the informed consent form.
For practical and logistical reasons, invitations for the nationwide study (and thus for the current study) were sent out consecutively in batches consisting of invitations to 30 to 50 women. The calculation of the target sample size was based

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 109

on the expected proportions of participants in both randomization groups filling out the paper-based questionnaire. Based on a previous study by Quigley et al. [217], in which similar randomization groups were used, it was estimated that 77% of participants in the mixed invitation group and 27% of the participants in the Web-only invitation group would complete and return the paper-based version of the questionnaire. With 95% power and a significance level of .05, it was estimated that a minimum of 26 participants would be required in each group [218]. However, it was decided to include all women who were invited for the nationwide study during a fixed period of time (ie, January 1, 2009, through May 31, 2010), thereby assuring that the target sample size would be met.

Randomization occurred by sorting the survivors alphabetically based on the street name of their address, after which the first half of the survivors was allocated to the mixed invitation group and the second half to the Web-only invitation group.

All eligible female survivors received a study information package by postal mail consisting of an informed consent form, a refusal form, a postage-paid reply envelope, and an instruction sheet with personalized log-in details. Depending on the allocated randomization group, a paper-based questionnaire was added to this study information package. The envelope containing the study information package was sealed and put in another envelop together with a cover letter, signed by the head of the relevant pediatric oncology department, in which the study was explained very briefly. This was done in order to give survivors the chance to return the entire study information package without having to open the envelope containing this package and without having to read the extensive study information. Thus, survivors could respond to the study information package that was sent in four different ways. These were: (1) sending back a filled-out questionnaire (either Web-based or paper-based) together with a filled-out informed consent form, (2) sending back a filled-out informed consent form only in cases where the potential participant was not willing to fill out the questionnaire but was willing to take part in other parts of the study, (3) sending back a filled-out refusal form, (4) sending back the entire study information package marked return to sender. For the purpose of this study, survivors were categorized as being responders if they chose one of the four above-mentioned response options, otherwise they were categorized as nonresponders.

Participants in both groups were assured that all information provided both by the paper-based as well as the Web-based questionnaire was confidential. Moreover, it was mentioned that data provided via the Web-based version were transmitted over a secure Internet connection and could not be viewed by unauthorized persons.

Follow up and reminders

If an envelope appeared undeliverable because of an incorrect or nonexistent postal address, the online telephone directory was used to find the correct address. If this proved unhelpful, vital status and current address were checked by means of the Gemeentelijke Basis Administratie (Dutch Municipal Population Register).

If the questionnaire was not returned within 3 weeks, a reminder was sent by postal mail. For participants in the mixed invitation group, this reminder consisted of

a letter in which the relevance of the study was again stressed and in which the individual was asked to respond. For participants in the Web-only invitation group, a paper-based version of the questionnaire was added to this reminder letter. When, after 3 weeks, still no response was received, patients in both groups were contacted by telephone and were asked to respond.
For the purpose of the current study, response time is defined as the time (number of days) elapsed between the day the envelope with the study information package was sent and the day a response was received.

Data analysis

Data were analyzed using SPSS for Windows, version 15.0 (SPSS Inc, Chicago, IL, USA). Descriptive statistics were used to describe differences between (1) participants allocated to the mixed invitation group and the Web-only invitation group, (2) respondents filling out the paper-based questionnaire and the Web-based questionnaire, and (3) responders and nonresponders. Independent samples t tests and Pearson chi-square tests were used to test whether these differences were statistically significant. A P value of less than 0.05 was considered to be statistically significant.
Multivariable logistic regression analysis was used to predict the probability of filling out the Web-based version of the questionnaire by reporting odds ratios (ORs) and 95% confidence intervals (CIs). A prediction model was developed using a backward selection procedure with a P value of 0.10 as the criterion for exclusion of variables.

RESULTS

General response characteristics

Included in this study were 277 women. Table 1 outlines the response characteristics of the participants allocated to the two randomization groups.
Response rates were comparable in both groups (83% in mixed invitation group vs 89% in Web-only invitation group, P = 0.20). In addition, participation rates—defined as the number of women who were willing to fill out the questionnaire—did not differ significantly between the mixed invitation group and the Web-only invitation group (66% or 90/137 vs 59% or 83/140, respectively, P = 0.27). Moreover, median response time was comparable in both groups. In the mixed invitation group, significantly more respondents filled out the paper-based questionnaire compared with the Web-only invitation group (83% or 75/90 and 65% or 54/83, respectively, P = 0.01).
Overall there were no differences between the responders in the mixed invitation group and the Web-only invitation group with respect to the number of women responding after the initial invitation, after the first reminder, and after the second reminder. However, when the results regarding the timing of the response are related to the type of questionnaire filled out by the respondents, some differences between the two groups appear. Among the group of responders who sent back either type of the questionnaire before the first postal reminder was sent, that is, the "fast responders," the proportion of responders filling out the paper-based version of

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 111

Table 1 Response characteristics of the participants receiving the mixed invitation and the web-only invitation (n=277)

	Mixed invitation group (n=137)	Web-only invitation group (n=140)	P value
Number of responders, n (%)	114 (83)	124 (89)	0.20
Response time, days; median (IQR)	32.5 (54.5)	34.5 (44.8)	0.51
Reminders sent, n (%)			
By mail	88 (64)	93 (66)	0.70
By telephone	29 (21)	31 (22)	0.84
Timing of response, n (%)			
After initial invitation	54 (47)	53 (43)	
After 1st reminder (by mail)	42 (37)	45 (36)	0.56
After 2nd reminder (by telephone)	18 (16)	26 (21)	
Type of response, n (%)			
Returned envelope to sender	6 (5)	10 (8)	
Refused (sent back refusal form)	18 (16)	31 (25)	0.12
Willing to participate	90 (79)	83 (67)	
Type of questionnaire filled out, n (%)			
Paper-based	75 (83)	54 (65)	0.01
Web-based	15 (17)	29 (35)	

the questionnaire was significantly larger in the mixed invitation compared with the Web-only invitation group (74% or 32/43 vs 51% or 18/35, respectively, P = 0.04). This difference in type of response remained after the first postal reminder (to which a paper-based version of the questionnaire was added) was sent, but it was no longer statistically significant: 91% (30/33) in the mixed invitation group compared with 75% (24/32) in the Web-only invitation group filled out the paper-based version of the questionnaire after being reminded by postal mail (P = 0.09).

When the 238 women who responded to the study invitation by sending back the questionnaire, the informed consent form, the refusal form, or the entire study information package were compared with the 39 women who did not respond at all, it appeared that these two groups did not significantly differ regarding age, age at diagnosis, or type of diagnosis. The nonresponder group included 20 survivors (12 of 137 or 9% in the mixed invitation group and 8 of 140 or 6% in the Web-only invitation group) whose postal address could not be verified and who could not be contacted by telephone either. It was decided to consider these survivors to be nonresponders. However, they might not be “true” nonresponders since it is not known whether they indeed received the study information package and the postal reminder.

Comparing the 173 women who participated in this study with the 104 women who did not participate (ie, women indicating they refused to participate and women who did not respond) did not reveal significant differences regarding current age or

age at diagnosis. However, it appeared that the proportion of women with leukemia was significantly higher in the participant group compared with the nonparticipant group (52% or 90/173 vs 38% or 39/104, respectively, P = 0.02) (data not shown).

Characteristics of questionnaire respondents

Figure 1 summarizes the participant flow.

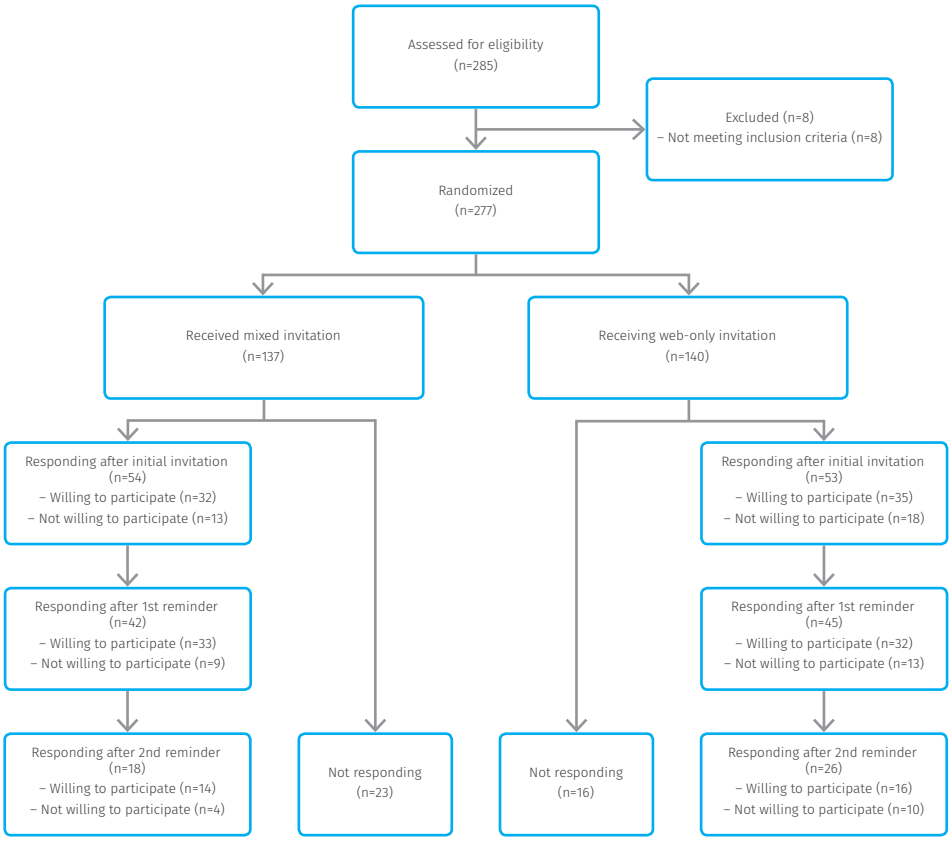


Figure 1 Flow diagram of the participants.

It appeared that the 44 women who filled out the Web-based version of the questionnaire were more likely to have had a higher educational level than the 129 women who filled out the paper-based version (P = 0.01). No differences were found regarding age, type of diagnosis, age at diagnosis, employment status, or marital status. Among the group of respondents filling out the questionnaire, it was investigated which factors influenced the probability of filling out either the paper- or the Web-based version of the questionnaire. Table 3 shows the odds ratios and 95% confidence intervals (CI) for the variables in the final model of the logistic regression analysis. Age, educational level, employment status, and randomization group were significant factors influencing the probability of filling out the Web-based questionnaire.

Table 2 shows the characteristics of the 173 women who returned a questionnaire.

Table 2 Characteristics of respondents filling out the paper-based and the web-based questionnaire (n=173)

	Paper-based questionnaire (n=129)	Web-based questionnaire (n=44)	P value
Age, years; mean ± SD (range)	29.7 ± 7.9 (18.8-52.3)	30.9 ± 8.6 (19.4-52.1)	0.40
Type of diagnosis, n (%)			
Leukemias	70 (54)	20 (45)	
Lymphomas	17 (13)	9 (21)	
Brain and central nervous system cancers	5 (4)	3 (7)	
Bone tumors	10 (8)	2 (5)	0.73
Neuroblastomas	6 (5)	1 (2)	
Germ cell tumors	3 (2)	2 (5)	
Nephroblastomas	5 (4)	3 (7)	
Other	13 (10)	4 (9)	
Age at diagnosis, years; mean ± SD (range)	7.4 ± 4.7 (0.4-19.5)	8.9 ± 4.6 (0.6-15.9)	0.07
Educational level, n (%) [a]			
High	30 (24)	19 (43)	
Medium	78 (62)	24 (55)	0.01
Low	18 (14)	1 (2)	
Employment status, n (%)			
Unemployed	28 (22)	4 (9)	
Student	15 (12)	8 (18)	0.12
Employed	82 (66)	32 (73)	
Marital status, n (%)			
Never married	86 (67)	29 (66)	
Married	39 (31)	15 (34)	0.56
Divorced	3 (2)	0	

[a] *Categorized as low: up to and including lower technical and vocational training; medium: up to and including secondary technical and vocational training; high: up to and including higher technical and vocational training and university*

More specifically, the probability of filling out the Web-based questionnaire was higher for participants allocated to the Web-based invitation group for participants who were older, and for participants with a higher educational level. Finally, students appeared to have a higher probability of filling out the Web-based questionnaire compared with participants who were employed.

Table 3 Factors associated with the probability of filling out the web-based version of the questionnaire: results of logistic regression

	P value	OR	95% CI	
			Lower	Upper
Age	0.01	1.08	1.02	1.15
Educational level (ref. group: High level)	0.04			
Medium		0.65	0.28	1.53
Low		0.06	0.01	0.52
Employment status (ref. group: Employed)	0.03			
Student		3.25	1.00	10.56
Unemployed		0.35	0.10	1.29
Randomization group (ref. group: Mixed invitation grp)	0.01	2.85	1.31	6.21
Nagelkerke pseudo R^2		0.21		

DISCUSSION

Statement of principal findings

In the present study, we examined differences in response between female childhood cancer survivors who received either a mixed invitation (paper-based questionnaire together with log-in details for Web-based questionnaire) or a Web-only invitation (log-in details only). The results show that survivors receiving the mixed invitation preferred filling out the paper-based version instead of the Web-based questionnaire as compared with the survivors receiving the Web-only invitation. Thus, when a paper-based version of the questionnaire was added to an invitation in which also the possibility of filling out the Web-based version was mentioned, the survivors were more likely to choose the paper-based option. Moreover, when the results regarding the timing of the response are taken into account this finding is endorsed since a large proportion (75%, 24/32) of females who initially received the log-in details only responded by filling out the paper-based questionnaire after they received a postal reminder (3 weeks later) to which a paper-based version of the questionnaire was added. This proportion is comparable to the proportion of females filling out the paper-based questionnaire immediately after the invitation (ie, before the postal reminder) among those who initially received the login details together with the paper-based version of the questionnaire (75%, 24/32).

Comparison with other studies

To our knowledge, no studies are available that have compared response rates to a Web- and paper-based version of a questionnaire on reproductive and fertility issues among young adult women. However, a few studies are available that have evaluated these issues by means of a Web-based questionnaire only. In a group of female survivors of breast cancer, the response rate to this type of questionnaire was 51% [219 220] whereas in a group of women aged 17 to 21 years, this rate was 72% [221]. However, no information was provided regarding characteristics of the nonresponder group.

In our study, the overall response rates in the mixed invitation group and the Web-only invitation group did not differ. This result is in line with the results found in the study of Quigley et al. [217]. In the study by Quigley et al., military personnel were requested to participate in a survey on information services. In one study group, a paper-based questionnaire was used with an added option of completing the questionnaire via the Internet. In the other study group, an Internet-based questionnaire was used with an added option of completing a paper version of the questionnaire by mail. Although response rates in both study groups were lower (42% and 37% respectively) than the response rates found in our study, differences in response rates between the two groups were not found, as was the case in our study. Furthermore, of the participants receiving the paper-based questionnaire with the Internet option, 77% chose to complete paper-based questionnaire. In our study, a similar proportion of participants in the mixed invitation group filled out the paper-based questionnaire, that is, 83%. However, Quigley et al. found that of the participants receiving the Web-based questionnaire with the option of the paper-based version, 73% chose to complete the Web-based questionnaire, while in our study the proportion of women in the Web-only invitation group who filled out the Web-based questionnaire was much lower (35%).

Furthermore, the participation rates (ie, the proportion of women who filled out the questionnaire) measured in our study can be considered as being rather high (66% in the mixed invitation group and 59% in the Web-only invitation group). In other studies using Web-based questionnaires in combination with paper-based versions these rates are, in general, lower [209 217 222]. A possible explanation for the high response rates found in our study might be the salience of the study topic. It is known that potential participants are more likely to respond to both paper-based and Web-based surveys when the salience of the topic, defined as the degree to which the topic is of interest or is relevant for participants, is high [195 202 223]. Moreover, the questionnaire used in the current study was one of the three study components used in a nationwide study on fertility issues in female childhood cancer survivors, with the other two study components being the provision of a blood sample and a transvaginal ultrasound measurement of the reproductive organs. It is known that female survivors of childhood cancer are in need of information regarding their reproductive function [224 225]. Therefore, participation in this study might be appealing for a large group of the invited females, resulting in higher response rates compared with studies in which a questionnaire is the only measurement instrument used.

Our results show that the use of reminders improved the response rates substantially. After the first reminder (a letter sent by postal mail), the response almost doubled in both randomization groups. Other studies support our finding that both postal and telephone reminders are effective in increasing response rates for both Web-based surveys as well as traditional paper-based surveys [206 226 227].

In our study, the majority of the respondents preferred filling out the paper-based version of the questionnaire to filling out the Web-based version. Moreover, age, educational level, and employment status appeared to be important factors influencing the decision to fill out either version of the questionnaire. Our finding that women with a high education level as well as students tended to choose Web-

based questionnaires over paper-based questionnaires is in line with previously published results [195 222 228]. However, results of the present study could not endorse other study results stating that Web-based questionnaires are more likely to attract younger respondents than paper-based questionnaires [198 228].
Another factor that may have played a role in the decision of the respondents to fill out either the paper-based or the Web-based version of the questionnaire is the length of the questionnaire used in the current study, which was rather long. The paper-based version consisted of 122 questions covering 32 pages. The Web-based questionnaire required several computer screens, with the number of questions on one screen depending on the type and length of the questions. For women filling out the Web-based version of the questionnaire, the median (IQR) time spent on filling out the questionnaire, which was automatically registered by the Web-based questionnaire tool, was 42.7 minutes (28.7 minutes to 67.8 minutes). Unfortunately, in the current study these data were not collected for the group of women filling out the paper-based version of the questionnaire. However, in the larger nationwide study, of which this study is part, a question was added to the paper-based questionnaire at a later point in time asking how much time was spent filling out the questionnaire. Median (IQR) time spent in this group (n = 145) was 30.0 minutes (30.0 minutes to 60.0 minutes) minutes. Thus, although this information was recorded among a different group of participants, it seems that filling out the paper-based version of the questionnaire took less time compared with the Web-based version. Various studies have shown the length of both Web-based and paper-based questionnaires to be negatively related to response rates [223 229-231]. However, no literature is available on differences in response rates to paper-based and Web-based questionnaires related to the length of a questionnaire. The results of our study seem to indicate that people tend to choose the paper-based version of a questionnaire when it concerns a long questionnaire. However, whether shorter surveys result in higher response rates when they are offered through the Web and longer surveys result in higher response rates when offered on paper needs further investigation.
In addition, the topic of our questionnaire can be considered to be rather personal. It is known that questionnaires containing questions of a sensitive nature result in lower response rates [223 229]. Although to our knowledge, studies investigating differences in response rates to paper-based and Web-based questionnaires taking into account the degree of sensitivity of the questions are lacking, one could assume that the sensitivity of our topic may have resulted in more respondents filling out the paper-based questionnaire especially since it is known that respondents filling out questionnaires through the Internet have doubts about their privacy and the confidentiality of their responses [204 232]. Despite the fact that data security and confidentiality were stressed in the letter accompanying our Web-based questionnaire, this might have led to more women filling out the paper-based questionnaire.

Limitations of current study

An important limitation of our study is the generalizability of the results found. Our study population mainly consisted of relatively young women, and thus the results may be less representative of older age groups or mixed groups including males.

Moreover, our study population represents a rather unique clinic population, that is, long-term survivors of childhood cancer. In addition, the topic of the questionnaire used cannot be considered a conventional subject. Therefore, caution should be exercised when translating the results found in the current study to other study groups or other study topics. Furthermore, the available data on the nonresponders in the present study were limited to age, age at diagnosis, and type of diagnosis. As a consequence, potential bias introduced due to nonresponse, also influencing the generalizability, could not be investigated extensively. However, in many of the studies using both paper-based as well as Web-based questionnaires, data on nonresponders are not available at all. As nonresponse to surveys seems to be increasing in recent years [233 234], future studies investigating the degree of bias as well as its consequences for the interpretation of data collected by paper-based and Web-based questionnaires are of great importance.

CONCLUSIONS

Survivors of childhood cancer from this era represent a highly mobile group, and they may not be as available or as responsive to contact by traditional mail methods [235]. Successful recruitment of this population will require new methods of contact such as email and Web-based methods. Therefore, although our findings indicate that most survivors preferred the paper-based version over the Web-based version when offered both, we conclude that Web-based questionnaires are promising data collection tools for childhood cancer survivors. However, researchers should carefully weigh the methodological benefits and barriers of using either a paper-based or a Web-based questionnaire for this group of subjects, taking into account possible response bias.

516278-L-bw-Overbeek
Processed on: 19-1-2018

7

Intra-cycle fluctuations of anti-Müllerian hormone in normal women with a regular cycle: a re-analysis

A. Overbeek
F.J. Broekmans
W.J. Hehenkamp
M.E. Wijdeveld
J. van Disseldorp
E. van Dulmen-den Broeder
C.B. Lambalk

Reprod Biomed Online.
2012 Jun;24(6):664-9.

ABSTRACT

Anti-Müllerian hormone (AMH) has emerged as an important marker of ovarian reserve. Its variation throughout the cycle, however, may still be a matter of debate. The objective of this study was to re-evaluate the intra-cycle fluctuations of AMH in individuals in a prospective clinical study with focus on the age-related effects on these fluctuations. Frequent blood samples were obtained from the mid-luteal phase of the first cycle to the mid-luteal phase of the second cycle in 44 healthy, regularly menstruating Caucasian women. Main outcome measures were individual fluctuations of AMH concentrations during the natural menstrual cycle. AMH concentrations exhibited large fluctuations throughout the cycle and did not follow a defined pattern. Female age was negatively correlated with mean AMH concentrations. The absolute intra-individual variation was also negatively associated with age, whereas the relative intra-individual variation was positively associated with age. Although the fluctuation in relative intra-individual variation was higher in the older group, the absolute variation is very low and these fluctuations might therefore be of limited clinical relevance in this age group. These data show that in younger women caution should be exerted with the interpretation of a single randomly taken AMH measurement as a representative of ovarian reserve.

INTRODUCTION

Anti-Müllerian hormone (AMH) is a member of the transforming growth factor superfamily and is produced in granulosa cells in the primary and pre-antral follicles of the ovary [236]. It plays a role in inhibiting follicular recruitment and FSH-dependent growth and may regulate the selection of pre-antral and small antral follicles [14-16].

Various methods of measuring AMH are available. Historically, AMH concentrations have been measured by either the Immunotech (IOT) assay (Beckman Coulter, Webster, Texas, USA) or the Diagnostic System Laboratories assay (DSL; Diagnostic System Laboratories, Webster, Texas, USA). While several studies have demonstrated wide variations in the results within each of these assays [237-240], there does appear to be excellent correlation between the two methods [238]. Recently, a second generation AMH enzyme-linked immunosorbent assay (ELISA), named AMH gen II, has been developed to measure AMH in serum; this has been standardized to the IOT assay [241].

AMH concentrations vary with age; they are low at birth and increase at puberty, reaching a maximum value in late puberty and then decreasing to undetectable values at menopause [242]. Concentrations of AMH may also be influenced by other variables. There is some evidence, for example, that they vary with race [243]. Some studies [244-246], but not all [247 248] have also found a correlation between body mass index and AMH concentrations.

Research evaluating fluctuations of AMH concentrations within the menstrual cycle is contradictory. La Marca et al. [21] concluded that serum AMH concentrations did not change significantly during the menstrual cycle. This finding was confirmed by another group [22] and Hehenkamp et al. [239] concluded that AMH concentrations did not show consistent fluctuation patterns like FSH, LH and estradiol, but did show small random fluctuations. In contrast, several publications have noted significant fluctuations within a cycle in humans and primates [240 249-253]. However, it remains debatable whether these fluctuations are clinically relevant and there are other possible limitations associated with these studies. For example, it was unclear which assay was used to measure AMH in the study by Cook et al. (2000) [250], the results of Lahlou et al. (2006) [251] were reported in a congress abstract and never published as a full article, Wunder et al. (2008) suggested that fluctuations might have been seen because they measured AMH daily and had a significantly younger study population than other studies [253] and Streuli et al. (2009) noted that the fluctuations they measured were smaller than the intercycle variability and were therefore not clinically relevant [240].

AMH patterns during the cycle have recently been categorized into two groups: the ageing ovary pattern and the younger ovary pattern. In the ageing ovary pattern, concentrations of AMH were low and there was little fluctuation, while AMH concentrations were higher in the younger ovary pattern, with significant fluctuation during the follicular phase [254]. However, it appears that the statistical analyses in this study were flawed, as groups were categorized according to AMH concentrations despite AMH itself being the main outcome of the study [255]. Additional research in which statistical analyses are properly conducted are needed to shed light on the

hypothetical relationship between AMH fluctuation and age. Nevertheless, the study by Sowers et al. clearly illustrated the influence of age on AMH concentrations [254]. Regardless of the debate on the fluctuations of AMH within the menstrual cycle, several studies have demonstrated a correlation between AMH concentrations and response to assisted reproductive techniques. Recently, it was shown that a single random measurement of AMH can predict poor response with relatively good specificity (reviewed by Broer et al. [256]). It has also been reported that the mean intracycle variability in serum AMH concentrations is less than that associated with antral follicle counts, indicating that this could be a robust means of assessing ovarian reserve in subfertile women [257]. It is therefore expected that, in the near future, AMH concentrations will become an increasingly important factor in ovarian reserve testing and clinical decision-making.
However, the conclusions of these studies are based on the mean values of the study population and do not take into consideration the large random fluctuations that can occur when evaluating an individual's AMH plot throughout the cycle. It is therefore essential to evaluate individual AMH fluctuations to test whether taking a single untimed sample could be justified as a means of assessing ovarian reserve. Consequently, this study has re-evaluated the data previously described in the study by Hehenkamp et al. (2006) to assess the effect of female age on the degree of fluctuation of AMH in each separate individual [239]. This study also tested the ageing ovary hypothesis [254] by comparing AMH fluctuations in two groups of women aged ≤ 38 and > 38 years.

METHODS

Patients and study design

The complete data associated with a study population previously described by Hehenkamp et al. (2006) were re-evaluated to analyse possible relevant fluctuations in AMH concentrations during the menstrual cycle. Approval was obtained from the local ethics committee and written informed consent was obtained from all participants. In brief, 44 healthy Caucasian women were recruited by advertisements in local newspapers. Participants were required to meet the following criteria: regular ovulatory menstrual cycle, proven natural fertility with spontaneously arising pregnancy within 1 year after the start of unprotected intercourse, no history of ovarian disease or surgery and no use of hormonal contraception. Measurements were started in the mid-luteal phase of the first study cycle. From the seventh day following a temperature rise on the basal body temperature chart, the subjects visited the clinic every 2 or 3 days for blood sampling until the occurrence of menstruation. The volunteers returned on cycle day 2, 3 or 4 after the onset of menstrual bleeding in the second study cycle and every 2–3 days thereafter for blood sampling and ultrasonography to observe the development of the dominant follicle, until this follicle had reached a diameter of at least 14 mm. Ultrasonography was repeated daily until 4 days after ovulation. The day of ovulation was defined as the day that the follicle completely disappeared or at which a reduction of at least 5 mm in the average diameter was seen. AMH concentrations were measured

in serum. Specimens were stored at –20°C until processed. AMH concentrations were estimated using the DSL enzyme-immunometric assay. Inter- and intra-assay coefficients of variation were less than 5% at the concentration of 3 µg/l and less than 11% at the concentration of 13 µg/l. The lower detection limit of the assay was 0.026 µg/l.

Statistical analysis

Statistical analysis and construction of graphs were carried out using the Statistical Package for Social Sciences version 14.0 (SPSS, Chicago, Illinois, USA). The mean AMH concentration was calculated for each individual subject. To evaluate the intra-individual changes, the absolute maximal difference between AMH concentrations per individual (absΔAMH) was determined, by subtracting the lowest AMH value in one cycle from the highest AMH value in that cycle, for each individual. Curve estimations were carried out to evaluate which model best fitted the relationship between age and absΔAMH. Correlations were evaluated between mean AMH concentration per cycle and body mass index, and mean AMH concentration per cycle and age, using Pearson's correlation tests and linear regression analyses. The predictability of a single AMH value was measured by dividing the patient population into two groups, according to age: (i) on or under the median (≤ 38 years) and (ii) above the median (>38 years). The expected drop in AMH when age increased by 5 years was calculated using a regression equation. Afterwards, it was calculated how often the absolute ΔAMH would be greater than the expected 5-year decline.

RESULTS

The basic characteristics of the 44 women recruited for this study (median age 38.3 years, range 25.6–46.2 years) are summarized in Table 1.

Table 1 Basic characteristics of healthy female volunteers (n=44)

Basic characteristics	
Age (years)	38.3 (25.6 – 46.2)
BMI (kg/m²)	24.9 ± 5.2
AMH level (µg/L)	1.05 ± 0.92
Absolute difference in AMH (absΔAMH, µg/L)	0.56 ± 0.53

Data are presented as medians (range) or means ± SD. Absolute difference in AMH: highest minus lowest AMH level per individual.

A total of 396 blood samples was collected (median per volunteer 9, range 5–14). Figure 1 shows all the individual measurements. The overall AMH concentration was (mean ± SD) 1.05 ± 0.92 µg/l (range 0.01–4.60 µg/l). The absΔAMH was 0.56 ± 0.53 µg/l (range 0.03–2.37 µg/l).

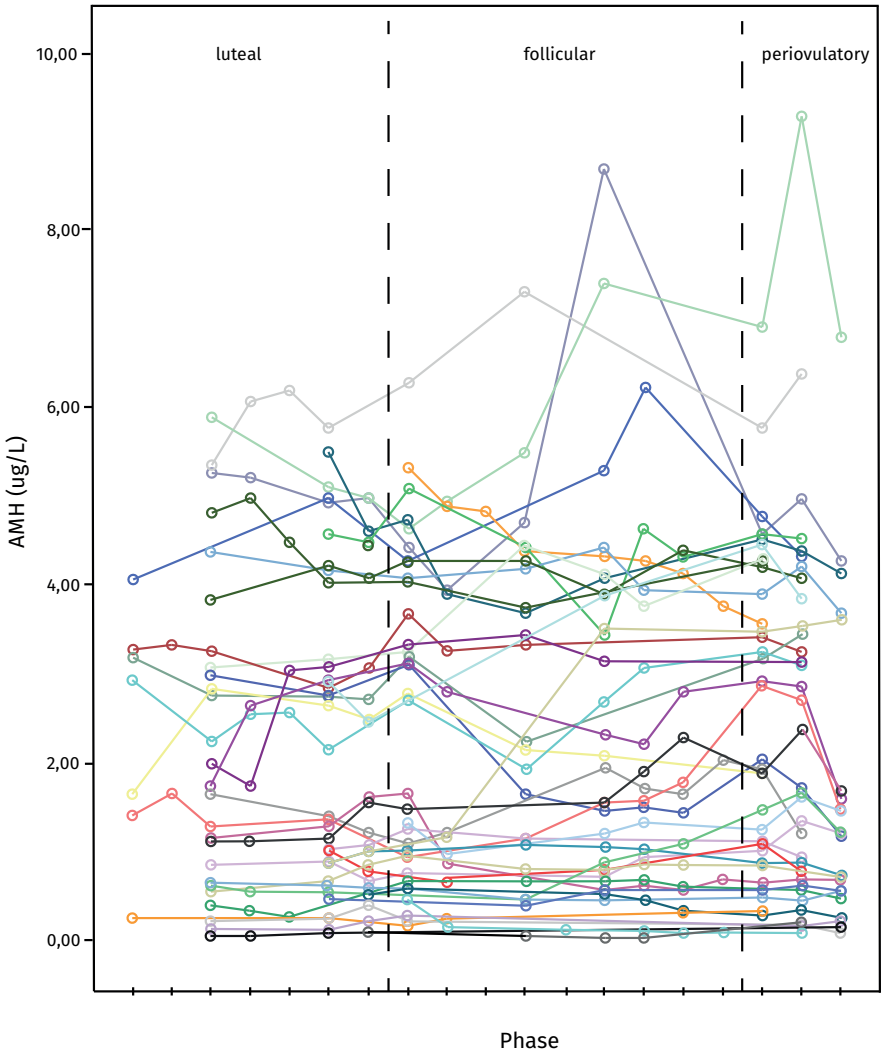


Figure 1 Individual patient plots of AMH during the cycle. Each circle represents a sample. The dashed lines depict the different phases in the cycle.

Age was significantly negatively correlated with mean AMH concentrations (R^2 = 0.64, P < 0.001) (Figure 2, left). A linear model fitted the data best. Inserting a quadratic or logarithmic term did not improve the model. A regression equation was calculated from the linear correlation: AMH = 4.70 + (–0.10 × age). No significant correlation was found between body mass index and the mean AMH concentration per cycle (Figure 2, right).

Absolute intra-individual variation (absΔAMH) was significantly negatively associated with age: younger women had significantly larger changes in absΔAMH than older women (R^2 = 0.40, P < 0.001) (Figure 3).

516278-L-bw-Overbeek
Processed on: 19-1-2018

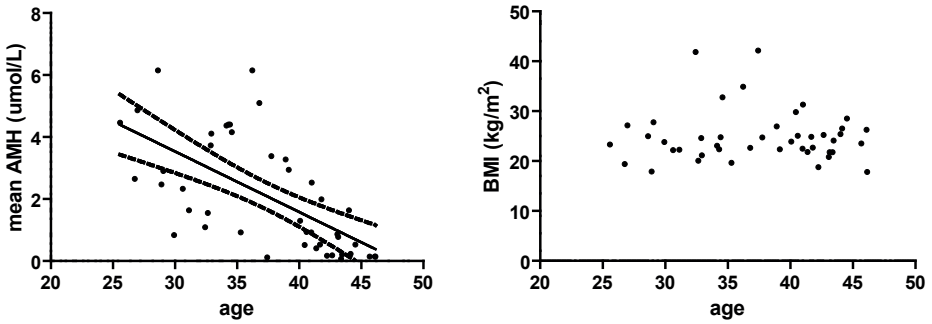


Figure 2 Scatter plot depicting the relationship between anti-Müllerian hormone levels in serum and age (left) or BMI (right).

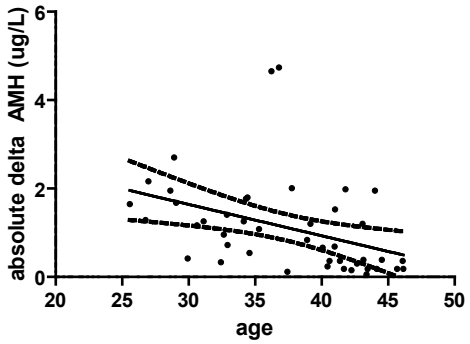


Figure 3 Scatter plot depicting the relationship between absolute intra-individual variation in anti-Müllerian hormone levels (absΔAMH) and age.

When calculated from the regression equation for the relationship between mean AMH concentration per cycle and age, the mean AMH concentration in the studied population (age range 25–46 years) can be estimated to drop by 0.5 µg/l over 5 years. Fluctuations were larger than 0.5 µg/l in one cycle in significantly more women in the younger age group than in the older age group (17/22 versus 6/22, P = 0.001) (Table 2).

Table 2 Comparison between age groups, divided according to median age

	Age		P value
	< 38 years (n = 22)	> 38 years (n = 22)	
Mean AMH (µg/L)	1.63 ± 0.87	1.64 ± 0.47	< 0.001
Absolute difference in AMH (µg/L)	0.81 ± 0.59	0.31 ± 0.29	0.001
Absolute difference in AMH > 0.5 µg/L (n)	17	5	0.001

Data are presented as means ± SD unless stated otherwise. Data were analyzed using ANOVA and the Chi-square test.

According to the above regression equation, this corresponds to a time span of 5 years. The absolute differences in AMH concentrations in the younger and older age groups, respectively, were 0.81 ± 0.59 µg/l and 0.31 ± 0.29 µg/l (P = 0.001).

DISCUSSION

Previous research has shown that AMH concentrations fluctuate randomly throughout the menstrual cycle with no relation to classical endocrine patterns. However, this re-analysis indicates that the amplitude of the fluctuations is larger in younger than in older women. This observation strengthens the hypothesis of an ageing ovary pattern to AMH fluctuations [254] and may have clinical implications.

It is known that AMH concentrations decline with age. However, in contrast to other studies, this study found that a linear relationship best fitted the AMH/age data. This was probably because of the small patient population and the age range of this population (25–46 years). Recent reports in larger populations show a quadratic relationship between AMH concentrations and age, where AMH concentrations are significantly higher in younger women [258 259]. This implies that the conclusions of the current study cannot be extrapolated to younger (< 25 years) or older (> 46 years) women and that additional research should be performed to evaluate the fluctuation of AMH in younger and older women. According to the regression equation, AMH concentrations will decline by 0.5 µg/l over 5 years, or 0.1 µg/l per year. This is comparable to the rate of decline described in an earlier longitudinal study [260]. In 17/22 women aged < 38 years (77%), AMH concentrations varied by > 0.5 µg/l within one cycle. Thus, within one cycle, differences may occur that equal the expected changes over a period of 5 years. This implies that the precision of AMH-based predictions in young women will be hampered to some extent by normal cycle fluctuation.

The difference between the current results, showing that AMH concentrations fluctuate throughout the cycle, and the results of older studies can be explained by the fact that each subject was evaluated individually. Hehenkamp et al. (2006) reported previously that while fluctuations of AMH concentrations are present, they are not substantial and do not consistently follow the known patterns of the classical hormones FSH, LH and estradiol. As knowledge of AMH fluctuations increased [254], the current study extended the analysis of Hehenkamp et al. (2006) using other statistical procedures and observed recognizable individual variations that appeared age-dependent.

The assay used for this study was the DSL assay. Because the newly available AMH Gen II assay was calibrated on the IOT assay, it will be difficult to directly compare these DSL assay results with any upcoming AMH gen II assay results. However, previous studies have shown that AMH concentrations measured using the IOT assay are consistently 2–4-fold higher than those using the DSL assay, resulting in an excellent correlation between the IOT and DSL methods [237 238]. Because of this, it is expected that similar intra-cycle fluctuations will be seen using either the IOT or the new method.

AMH concentrations are usually assessed to predict poor response to assisted reproductive techniques or as a marker of the extent of ovarian reserve at the end of a woman's reproductive life. If the absolute AMH concentration is low, a single measurement could be justified because the fluctuations will be of little clinical relevance. However, a single measurement might not suffice for predicting ovarian reserve in young patients with a high AMH concentration. Assessing AMH concentrations on more than one occasion or the use of ultrasound assessment of the antral follicle count (AFC) as an additional tool for predicting ovarian reserve in younger women is, therefore, recommended.

Nomograms for AMH concentrations are currently being developed [258 261] with the objective of facilitating the prediction of the commencement of menopause. Recent work from Sowers et al. has made it possible to calculate the age at which menopause will occur by means of a linear regression equation of logAMH [262]. Unfortunately, this equation has only been validated for older women (median age at the initial visit was 42 years) and can therefore not be extrapolated to a younger age group. Longitudinal data on AMH concentrations from adolescence to menopause are therefore necessary in order to adequately counsel patients about ovarian response or advancing menopause.

In summary, AMH concentrations can fluctuate substantially in younger women during the menstrual cycle. Overall, a single measurement of AMH appears to be highly correlated with the response to assisted reproductive techniques. However, large inter-individual variations in AMH concentrations mean that the accuracy of a single measurement as a marker for ovarian reserve could be unreliable.

516278-L-bw-Overbeek
Processed on: 19-1-2018

8

Between-method agreement and inter-rater reliability of antral follicle counts using conventional and three dimensional ultrasound in childhood cancer survivors

A. Overbeek
E. Wendel
M.H. van den Berg
F.E. van Leeuwen
J.W.R. Twisk
G.J.L. Kaspers
L.C. Kremer
M.M. van den Heuvel-Eibrink
W.J.E. Tissing
J.J. Loonen
A.B. Versluys
D. Bresters
E. van Dulmen-den Broeder
C.B. Lambalk
on behalf of the DCOG
LATER-VEVO study group

In preparation

ABSTRACT

The study's aim was to evaluate the inter-rater reliability and the between-method agreement of antral follicle counts (AFCs) using real-time 2D and stored 3D data in 50 childhood cancer survivors compared to 50 healthy sibling controls. Concordance correlation coefficients were 0.74 for inter-rater reliability and 0.66 for between-method agreement. Agreement was lower in subjects with higher AFCs and in overweight subjects. The inter-rater variability did not differ significantly between survivors and controls ($p=0.09$), however, the between-method reliability was significantly lower in the survivors group ($p=0.02$). Image quality was better in controls than in CCSs ($p=0.02$), with survivors being more frequently overweight ($p<0.001$). Application of uniform 3D ultrasound recording is favourable for multicenter study purposes allowing blinded single investigator data analysis, but the validity of counselling individual survivors using AFCs assessed in 3D mode as being representative of the ovarian reserve needs more research.

INTRODUCTION

Impaired fertility is one of the severe adverse effects women can face after having received chemotherapy and radiotherapy for a childhood malignancy. Chemo- and radiotherapy may substantially deplete the follicle pool, and since women are endowed with their entire non-replenishable follicle pool from birth, this can lead to a reduced ovarian reserve, infertility and/or premature menopause [50]. A thorough quantification of the remaining follicle pool may provide specialists with a tool to validly estimate the age at which menopause may occur. Transvaginal ultrasound assessing antral follicle counts (AFCs) may be such a tool. Because AFC shows a gradual decline with age, this may be a valuable marker to assess ovarian reserve in childhood cancer survivors at a young age, allowing them to make informed choices about if and when to try to conceive. AFCs have been shown to correspond well with histological analysis of the follicle pool. However, AFCs have been developed in an IVF-setting to predict ovarian response to hormonal stimulation [263]. As a consequence, studies were conducted in a subfertile population, and may not be comparable in order to assess ovarian reserve for childhood cancer survivors.
AFCs are calculated from ultrasound imaging with the conventional 2-dimensional (2D) technique or by the less commonly used 3-dimensional (3D) technique. The 3D ultrasound technique has several advantages. First, 3D ultrasound imaging is capable of visualizing all three orthogonal planes simultaneously. This allows the reconstruction of imaging planes that are not visible when using conventional 2D imaging, by using the stored volumetric data in combination with a computer program. These computer programs allow the assessor to automatically or manually measure the volume of each antral follicle. After all follicles have been identified in multiple planes, an image can be constructed in which only the follicles are shown in color (Figure 1).

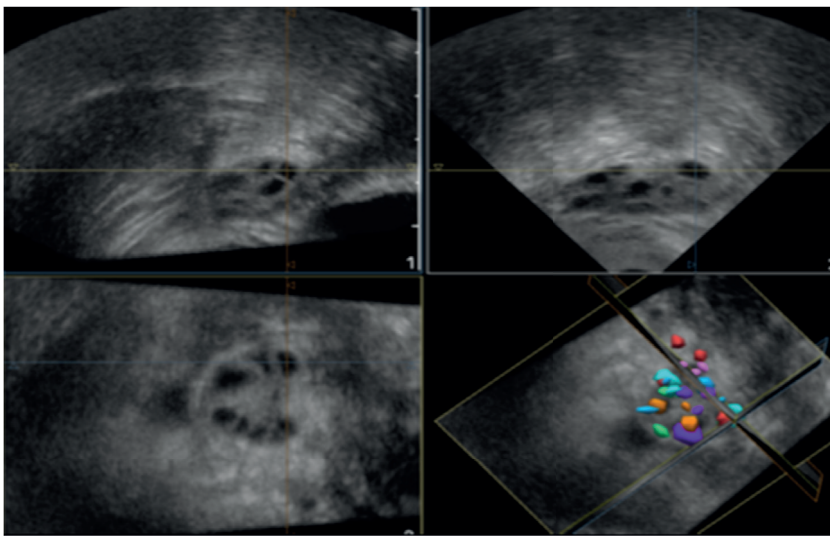


Figure 1 Images from an ovary in which follicles have been identified in all three orthogonal planes. Each follicle is given a specific color and is displayed in the lower right image.

For this study, these images and calculations were performed with custom-made Qlab 8.1 software (Philips Healthcare, Bothell, WA, USA). The second advantage is that the collection of the 3D ultrasound data can be done by different (inexperienced) persons without compromising data quality due to inter-rater reliability, as long as the assessment of the data is done by a single person according to earlier studies on inter-rater reliability [264 265]. 3D imaging techniques therefore may create new perspectives in assessing ovarian reserve and improve the methodological quality of ultrasound research. On the other hand, the offline evaluation of 3D data may take more time [264 265].

The inter-rater reliability of both 2D and 3D ultrasound measurements of antral follicle counts have been found to be high [264-266]. However, these data are limited in their practical use, as all studies that have been performed as yet, have been designed in such a way that the stored data have been re-assessed by two observers, rather than performing a completely new examination. In this way, additional sources causing variability in clinical practice (such as bowel movement or different handling of the ultrasound probe) are not taken into account [267]. The between-method agreement was found to be moderate and the degree of agreement between 2D and 3D AFCs seems to increase when higher numbers of follicles were counted [268]. Furthermore, Deb et al. found a lower antral follicle count in 3D-acquired data compared to a regular 2D AFC. It remains unclear whether this discrepancy in AFCs is caused by an underestimation of the follicles in 3D-mode or an overestimation of the follicles in 2D mode [269]. However, in a recent study comparing ovarian reserve tests with histologically determined primordial follicle numbers, it was found that AFC measured in 2D mode correlates very well with the actual primordial follicle numbers [29]. To our knowledge, no studies have been conducted in which AFC obtained by using the 3D-technique was validated in a histological model.

The reproducibility of antral follicle counts with 3D ultrasound has mainly been performed in women undergoing IVF/ICSI treatment. However, the 2D and 3D ultrasound imaging techniques have not been validated for females treated for cancer in the past. There are important differences between the subfertile patient population and childhood cancer survivors. First, the rationale for assessing antral follicle counts is different: in subfertile patients AFCs are often performed to predict ovarian response to assisted reproductive techniques. However, for childhood cancer survivors, counselling on their reproductive chances and the longevity of their fertile life span are very important. Second, former cancer treatment given to childhood cancer survivors might have induced changes in the reproductive organs, for example by radiation damage, which in turn might bring on changes in the reproducibility measures as well. It is well known that former cancer therapy causes premature menopause [39 40 270]. Histopathology of the ovary has shown that chemotherapy destroys the primordial follicles and can lead to ovarian atrophy in animal models. Other mechanisms causing damage to the ovary include injury to blood vessels and focal ovarian cortical fibrosis [50]. The current study was designed as a diagnostic study within an on-going Dutch nationwide cohort study (the DCOG LATER-VEVO study). The DCOG LATER-VEVO study was initiated in 2006, and aims to examine the effects of childhood cancer and its treatment on reproductive

function, ovarian reserve, and risk of premature menopause in female childhood cancer survivors [271]. The aim of the current study is to assess the between-method agreement of AFCs by comparing the counts made in real-time 2D to those obtained from stored 3D transvaginal ultrasound data. Additionally, we aimed to compare the inter-rater reliability by comparing the follicle counts obtained from 3D data by two different observers. Finally, we analysed whether the inter-rater reliability and the between-method agreement were different in survivors and controls, while controlling for confounders.

MATERIAL AND METHODS

Setting and participants

Eligible survivors for the DCOG LATER-VEVO study were randomly selected from a cohort of patients treated for childhood cancer at one of the seven Dutch pediatric oncology and stem cell transplant centres between 1963 and 2002. Within the collaborative Dutch Childhood Oncology Group (DCOG) an electronic database has been set up in each centre that includes patient and treatment details of all patients treated for cancer before the age of 18 years. The design of the nationwide study has been described earlier [271]. In short, fertility, pregnancy outcomes and menopausal status were addressed in a questionnaire, while ovarian reserve was assessed by a timed blood sample and a transvaginal ultrasound. The inclusion criteria for the nationwide study and the current study were identical and were defined as: female gender, having been treated for a malignancy or central nervous system tumor before the age of 18, having survived for more than 5 years, being alive and being at least 18 years before January 1st 2011. Female siblings, who were never diagnosed with cancer, who can read and speak Dutch, and who are 18 years or older, were approached as controls. For the current study, only subjects with a natural menstrual cycle (i.e. no hormonal contraceptives) between age 25 and 40 were included. Ultrasound measurements were performed on cycle day two to five. Subjects were excluded if they were not able to speak or read Dutch or if they were mentally retarded. Subjects were excluded when they had a history of pelvic surgery, when they reported to be amenorrheic or when they were found to have an ovarian cyst larger than 10 mm. In case the ovaries could not be visualized by means of the 2D ultrasound technique, the data of the subjects concerned were excluded. Approval for the nationwide study was obtained from the local ethics committee, and written informed consent was obtained from all participants.

Observers and ultrasound evaluation

Transvaginal ultrasound measurements of all individuals participating in the current study were performed by the same physician (AOV, 8 years experience with 2D ultrasound, 4 years with 3D ultrasound) using a HD11 XE ultrasound system with a 3D transvaginal probe (3 to 9 MHz; EnVisor HD, Philips Medical Systems, Eindhoven, the Netherlands). Initially, a 2D transvaginal ultrasound assessment of the reproductive

organs was performed to assess AFC and to determine length and width of both ovaries. An automated mechanical sweep with an angle of 90˚ of both ovaries was then acquired; data were stored and afterwards transferred to a personal computer for analysis. The AFC in each ovary was calculated from the number of antral follicles (size 2 - 10 mm) [264 272].
All 3D datasets were assessed once by two observers at least two months after the actual transvaginal scan using software to identify follicles (AOV and EWE). EWE had no prior ultrasound experience and was thus not experienced in recognizing ovaries on the screen. She received a short training in image recognition and counting antral follicles in the software program. Afterwards, several 3D datasets (not included in this study) were analysed by AOV and EWE and compared to ensure she was qualified to count antral follicles counts. All 3D data were analysed using custom-made Qlab 8.1 software (Philips Healthcare, Bothell, WA, USA) following a pre-specified protocol. The ovary was centred on the multiplanar view by using pan and zoom. Then, in GI 3D quantification mode, individual follicles were identified by using the stacked contour utility, which is a semi-automated contour utility. In the Slice Plane View all identified follicles could then be counted. No prior knowledge regarding the diagnosis or treatment was available to the assessors analysing the 3D data. The observers were blinded to the counts of the other observer (inter-rater reliability) or the 2D counts. The observers were aware that their judgment was used for the current study. Both observers performed the antral follicle counts using the 3D software once per individual subject. All analysis was performed using the same workstation by each observer, thus excluding bias caused by specific measurement settings.
The image quality of each dataset was subjectively categorized by one observer (AOV) into three groups, based on the proportion of the ovarian contour that could be seen clearly, according to Jayaprakasan [272]. Good quality was assigned when more than 90% of the ovarian contour was visible, median quality when 50-90% of the contour was visible and poor when less than 50% of the ovarian contour was visible.

Quantative variables

Basic characteristics of the included subjects were abstracted from the DCOG LATER-VEVO database (age), from the self-reported questionnaire (BMI) or from the blood samples collected for the nationwide study (FSH, qualitatively and quantitatively (proportion of women with FSH > 10 IU/L)). Blood samples were centrifuged for 10 min at 4°C (3000 rpm) within 30 min after venepuncture and frozen (-20°C) for storage until assayed. Laboratory screening was performed by the endocrine laboratory of the VU University Medical Center. Plasma FSH levels were analysed by an immunometric assay (Delfia, Wallac, Turku, Finland), with a lower detection limit of 0.5 IU/l. The intra and inter-assay coefficient of variation (CV) was 5 and 7%, respectively, at a concentration of 2 IU/l and 3 and 6%, respectively, at a concentration of more than 4 IU/l. Antral follicle were divided into three categories based on clinically relevant cut-offs (AFC < 5, AFC 5-11, AFC ≥ 12).

Study size

There is no standard program or statistical analysis available to calculate sample size needed for an agreement study. Liao et al. recommend a sample size of at least 32 [273], while Bland recommends a sample size of 100 [274]. At time of analysis, ultrasonographic data were available for 135 survivors of childhood cancer and 130 control subjects who fulfilled the inclusion criteria. A random selection of 50 survivors and 50 controls was taken.

Statistical methods

For the reproducibility analyses, each ovary was evaluated as a separate record. Obviously, both ovaries of one patient are related and, as such, cannot be seen as individual samples. However, in this study we only included patients of whom two ovaries were available for analysis. Therefore, even though the measurements are related, the data sample is balanced and will thus not influence the effect measure of the inter-rater reliability, but it will influence the standard error (which will be smaller due to the greater power).
A Wilcoxon rank test was used to test whether there were statistically significant differences in the antral follicle counts measured by two different observers and between different methods. In order to assess inter-rater reliability Lin's concordance correlation coefficients (CCCs) with confidence intervals were calculated, using a single measure, two-way mixed model [275]. To visualize the agreement among the 2D and 3D technique and among different observers Bland-Altman plots and limits of agreement were constructed from the relative differences. Weighted Cohen's Kappa was calculated after categorizing the number of follicles into categories based on clinically relevant cut-offs. Differences in basic characteristics between survivors and controls were analysed with an independent Student's t-test to compare the normally distributed data, and with Chi-square test to assess categorical data. A value of $P < 0.05$ was considered statistically significant. Additionally, Student's t-test was used to study differences in reproducibility measures of 2D and 3D between image quality and between survivors and controls. Analyses were performed using SPSS software (version 20.0, SPSS, Inc., Chicago, IL) and STATA software (version 13, StataCorp LP, College Station, Texas)

RESULTS

Participants

At time of analysis, ultrasonographic data were available for 135 survivors of childhood cancer and 130 control subjects who fulfilled the inclusion criteria. From these 265 subjects, seventy-four survivors and sixty-two controls were excluded because they were above 40 years of below 25 years of age, reported to have had a history of pelvic surgery, reported to be amenorrheic or were found to have an ovarian cyst larger than 10 mm. From those subjects who fitted the in- and exclusion criteria, a random selection of 50 survivors and 50 controls was taken. Figure 2 shows a flow chart of the patient accrual.

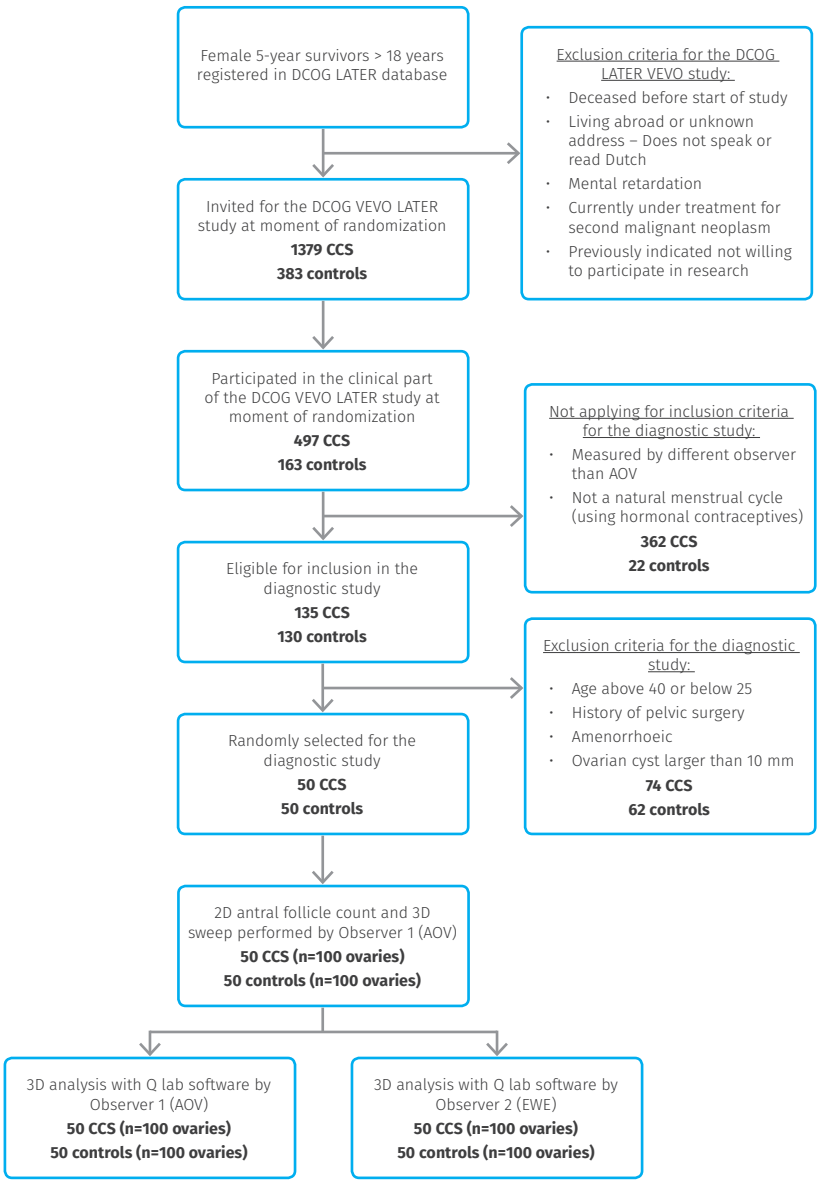


Figure 2 Flow chart of the accrual of participants for the 2D and 3D analysis.

Outcome data

In Table 1 the basic characteristics of the subjects are shown. The median age was 32.5 years (range, 24-49 years), median BMI was 22.5 (range 17.7-41.4) and the mean basal FSH value (SD) was 6.4 (2.1) IU/l. Three out of 100 subjects had an elevated FSH (cut off value 10 IU/l). The median (IQR) total antral follicle count (i.e. AFC of the left and right ovary combined) was 16 (14.3).

Table 1 General characteristics of the subjects

	Subjects (n=100)
Age (years, median, IQR)	32.5 (8.3)
20-25 years	10 (10%)
25-30 years	34 (34%)
30-35 years	29 (29%)
35-40 years	18 (18%)
40-45 years	8 (8%)
45-50 years	1 (1%)
BMI (kg/m^2, median, IQR)	22.5 (4.6)
Underweight (< 18.5 kg/m^2)	15 (15%)
Normal weight (18.5-25 kg/m^2)	55 (55%)
Overweight (25-30 kg/m^2)	20 (20%)
Obese class I (30-35 kg/m^2)	5 (5%)
Obese class II (35-40 kg/m^2)	3 (3%)
Obese class III (> 40 kg/m^2)	2 (2%)
Basal FSH (IU/l, median, range)*	6.4 (1.3-16.7)
25th percentile	5.1
50th percentile	6.4
75th percentile	7.4
FSH > 10 (n, %)*	3 (34%)
Total AFC (left and right ovary combined, measured with 2D technique) (median, range)	16 (2-60)
25th percentile	10.00
50th percentile	16.00
75th percentile	24.25

** FSH values were missing in 11 patients. Data analysed using Student's t-test for continuous data and Chi-square for categorical data.*

Main results

Inter-rater reliability

There was a significant difference in the median number of follicles counted by the different observers using the 3D method (observer 1, median 7 (range, 0-32); observer 2, median 6 (range 0-30), p < 0.001). Figure 3 depicts the Bland-Altman plot of the antral follicles counts with the 3D techniques between the two observers.
It seemed that with higher AFCs, there were larger differences between the two observers, with one observer systematically counting fewer follicles than the other. The concordance correlation coefficient was 0.74 (95% CI 0.69-0.80), which, according to the classification recently proposed by Martins and Nastri [267], was poor. Weighted Cohen's Kappa was 0.55 (0.46-0.65) indicating again only moderate agreement according to the classification of Landis and Koch [187].

Between-method agreement

The median number of follicles measured with the 2D technique was 8 (range, 0-32), whereas a median of 7 follicles (range 0-32) was measured by the same observer using the 3D technique. This was a significant difference (p = 0.03), indicating that in general higher antral follicles counts are measured with the 2D technique. The concordance correlation coefficient was 0.66 (95% CI 0.58-0.74). Figure 4 depicts the Bland-Altman plots of the antral follicle counts between the different techniques.

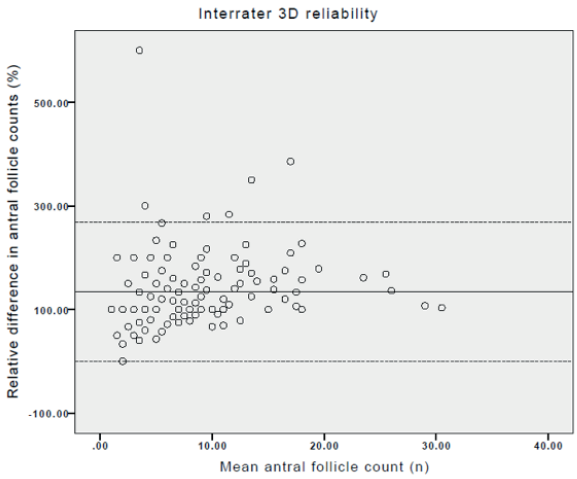


Figure 3 Bland Altman plot depicting lower and upper levels of agreement of inter-rater reliability.

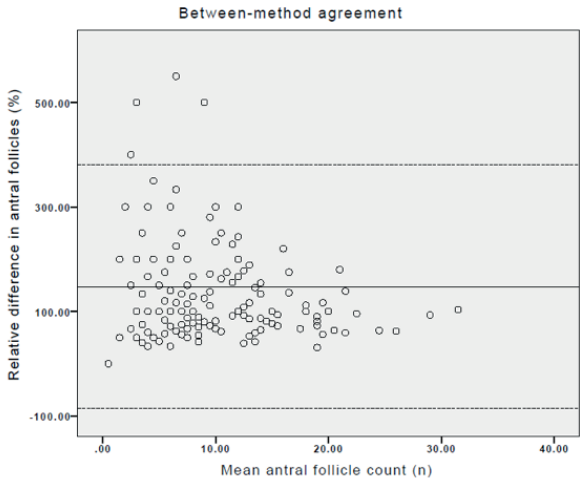


Figure 4 Bland Altman plot depicting lower and upper levels of agreement of between-method agreement.

Weighted Cohen's kappa was 0.52 (0.43-0.61), indicating a moderate agreement between categories of AFC measured with the 3D and the 2D technique

Childhood cancer survivors and control subjects

The median number of follicles counted in 2D for survivors was 8 (range 2-28) and 6 (0-30) for the controls. Survivors were significantly younger than controls (p<0.001) and had a higher BMI (p<0.001). Moreover, survivors were more likely to have a lower image quality than controls (35% vs. 19% respectively, p = 0.02) (Table 2).

Table 2 General characteristics comparing survivors and controls

Basic characteristics	Survivors	Controls	p-value
Age (years, median, IQR)	29 (24-42)	34 (25-49)	< 0.001
BMI (kg/m², median, IQR)	24.5 (18.0-41.4)	22.3 (17.7-40.8)	< 0.001
Basal FSH (IU/l, median, range)*	6.4 (1.3-11.9)	6.4 (2.9-16.7)	0.42
FSH > 10 (n, %)*	2 (2.3%)	4 (4.3%)	0.46
Image quality assessed in 3D mode (n, %)			0.02
Good	34 (34.7%)	32 (32.7%)	
Medium	30 (30.6%)	47 (48.0%)	
Poor	34 (34.7%)	19 (19.4%)	

FSH values were missing in 11 patients. Data analysed using Student's t-test for continuous data and Chi-square for categorical data

Although AFCs measured in 2D and 3D were comparable for controls (median number of follicles 8 in 3D mode vs. 7 in 2D mode, p=0.26), they were not for survivors (median number of follicles 6 in 3D mode vs. 8 in 2D mode, p<0.001). CCCs for inter-rater variability of controls and survivors both showed poor agreement (0.79 and 0.70 for inter-rater variability and 0.56 and 0.74 for between-method reliability for survivors and controls respectively) (Table 3).

Table 3 Outcomes for inter-rater reliability and between-method agreement comparing survivors and controls

	Subject	Relative mean difference	LOA-lower	LOA-upper	CCC	95% CI
Inter-rater	Survivor	127.51	-16.84	271.87	0.79	0.73-0.86
	Control	140.42	17.21	263.63	0.70	0.62-0.78
Between-method	Survivor	179.19	-102.10	460.47	0.56	0.43-0.69
	Control	117.29	-38.78	273.35	0.74	0.65-0.83

The inter-rater reliability was not significantly different between survivors and controls (p=0.09), however, the between-method. agreement was significantly lower in the survivor group (p=0.02). LOA; level of agreement, CCC; concordance correlation coefficient, CI; confidence interval.

Other analyses

Image quality

It was evaluated whether reproducibility characteristics were related to the image quality of the acquired 3D data. The median numbers of follicles counted in 2D and 3D for the highest image quality datasets were 10 (range 3-32) and 9 (range 2-32), respectively, which did not differ significantly. Also, for the medium image quality group, median follicle counts were comparable in 2D and 3D (median 7 (range 0-27) in the 2D group and 8 (range 1-27) in the 3D group). However, in the lowest image quality group, the median number of follicles counted in 2D differed significantly from those measured in 3D (median 6, range 2-28 in the 2D and median 4, range 0-17 in the 3D group, p <0.001). CCCs were calculated for inter-rater reliability according to image quality (Table 4). These data show that a lower image quality does not necessarily bring on larger inter-rater reliability.

Table 4 Outcomes for inter-rater reliability and between-method agreement when categorized in 3 subgroups based on image quality

	Quality	Relative mean difference	LOA-lower	LOA-upper	CCC	95% CI
Inter-rater	Good	124.67	29.21	220.12	0.77	0.68-0.85
	Medium	148.47	19.88	277.06	0.55	0.43-0.66
	Poor	127.51	-16.84	271.87	0.67	0.53-0.81
Between-method	Good	119.56	1.43	237.69	0.68	0.55-0.81
	Medium	119.22	-40.99	279.44	0.70	0.58-0.81
	Poor	233.41	-118.33	585.15	0.35	0.20-0.51

Certain confounders may influence image quality, such as age and BMI. Lower image quality was associated with a lower antral follicle count measured with the 2D method (ANOVA, p = 0.001), but not with age (p=0.51). High BMI was associated with lower image quality (p < 0.001). When BMI was categorized into overweight or normal weight (BMI< 25 or BMI ≥ 25), CCCs of the inter-rater reliability were comparable in both overweight and normal weight patients (0.74 vs. 0.73), but revealed a poor agreement (CCC < 0.90). However, the CCC of the between-method agreement was significantly lower in the overweight than the normal weight group (0.59 vs. 0.67). A CCC below 0.70 is associated with such a low agreement, that it is suggested that the method should not be considered for either clinical or research purposes. Numbers did not allow evaluating whether being underweight (BMI < 18, n=1) influenced the reproducibility.

DISCUSSION

Key results

We aimed to assess in childhood cancer survivors and controls the reproducibility of 3D ultrasonographic measurements of the antral follicle count, an important marker for ovarian reserve. We found that CCCs were 0.74 and 0.66 for inter-rater reliability

and between-method agreement, respectively, hence, indicating poor agreement. According to the recently published suggested cut-off values for interpretation of ICC (intra-class correlation) and CCC, this method might be employed for research purposes but should be considered with caution when being used in clinical practice because relevant differences in the method might occur only because of random errors [267].

Limitations

In our sample, CCCs were lower than ICCs reported in other studies [264 268]. The difference between our findings and previous study results may be due to the fact that different equipment (ultrasound equipment as well as software) was used or due to patient characteristics. For instance, for the current study, healthy controls as well as childhood cancer survivors were evaluated, whereas in the studies of Jayaprakasan, a subfertile study population was evaluated [264 272]. We speculate that former treatment (such as radiation damage) may influence image quality. In addition, a different statistical approach (CCC instead of average measures ICC) may have caused lower agreement levels. In this study, acquisitions were not independent. The acquisition is a potential source of variability and the estimates obtained examining the same acquisition will not represent the total variability that would be observed by repeating the exam. Therefore, reliability measures will tend to be overestimated [276].

Kappa values can only be calculated when continuous data are made categorical. We categorized our 2D counted antral follicle counts into three groups according to clinical cut-offs commonly used in literature (AFC < 5; poor; 5-11 normal; ≥ 12 high). A weighted Kappa takes into account that categories are ordered. Our study is the first to calculate weighted Kappa coefficients. It is therefore not possible to compare these outcomes to earlier findings. Our results showed only low to moderate agreement (0.55 and 0.52 for inter-rater reliability and between-method agreement, respectively). This indicates that AFCs estimated with the 3D technique often classified women more than one category higher or lower than when using the 2D technique. Although weighted Kappa values were low, the mean difference between the measurements was small, especially in those with a low AFC. This indicates that for research purposes in which groups are evaluated, both methods (2D and 3D) are suitable and that antral follicle counts can be reliably counted by different observers. However, when providing individual tailored advice to patients regarding fertility outcomes or pregnancy chances, one should keep in mind that qualification into low, normal or high ovarian reserve based on antral follicle counts alone may not be sufficient.

Overall, we observed that with the 3D technique in comparison to the 2D technique fewer follicles were counted. It remains controversial whether this difference is of clinical relevance (median of 8 vs. median of 7 follicles with the 2D and 3D technique, respectively). Deb et al. also found that fewer antral follicles were counted when applying post-processing software on 3D-acquired data compared to a regular 2D AFC [277]. On the other hand, Scheffer et al. counted more follicles in 3D mode as compared to the conventional 2D technique [268]. It remains unclear whether this discrepancy is caused by an underestimation of the follicles in 3D mode or an overestimation of the follicles in 2D mode [269].

Our data indicate that the higher the AFC values, the more likely AFCs will differ (either measured by two different persons, on two different time points or with two different techniques). When assessing inter-rater reliability, it appeared that observer 1 systematically counted more follicles than observer 2. Observer 2 was less experienced in analysing ultrasound data. Based on findings of other studies that showed excellent agreement between experienced and non-experienced assessors [272], it was not deemed necessary when designing this study to analyse the 3D data by two experienced assessors. However, as the current study suggests that there seems to be a non-random systematic misclassification, experience level may have influenced the results. However, it remains uncertain whether (in)experience caused the misclassification. Nevertheless, no direction of misclassification could be observed for between-method agreement.

When evaluating image quality, it seemed that in the lowest image quality group, median 2D antral follicle counts were higher than 3D AFCs, with the lowest CCC of 0.35 (0.20-0.51) in comparison to medium and good image quality (CCC 0.70 and 0.68 respectively). Image quality was associated with lower antral follicle counts and higher BMI, but not with older age. Although it is more likely that lower AFCs measured with the 3D technique are due to poor image quality (i.e. not being able to clearly see the follicles) rather than to an actual lower follicle count, this could not be confirmed by our study. Future studies evaluating the number of follicles numbers histologically and comparing these with antral follicle counts measured with the 3D ultrasound technique should shed some light on this issue. It might be very interesting to include patients who undergo a prophylactic oophorectomy (for example women with BRCA gene mutation) in such a study. Poor image quality was found significantly more often among survivors (34.7%) than controls (19.4%). This is also reflected in the relatively low between-method CCC of 0.56 (in comparison to 0.74 in controls). Histopathology of the ovary has shown that chemotherapy destroys the primordial follicles and can lead to ovarian atrophy in animal models. Other mechanisms causing damage to the ovary include injury to blood vessels and focal ovarian cortical fibrosis [50]. Radiotherapy to the pelvis may result in ovarian injury and diminished ovarian reserve, depending on the patient's age, treatment dose and irradiation dose [51]. This damage may influence the image quality. However, it is also very plausible that the lower image quality is due to the fact that the BMI of survivors was significantly higher than that of controls. Unfortunately, we were not able to evaluate the effect of treatment on the antral follicle counts and image quality, as these results have not yet all been gathered. We hope to be able to report on these factors in the near future.

In summary, this study shows poor CCCs for inter-rater and between-method reliability of the 3D technique as compared to the 2D technique for assessing antral follicle count. Obesity influences image quality and thereby reproducibility of 3D antral follicle counts. Image quality was lower in childhood cancer survivors, and it is unclear whether this is caused by a higher prevalence of obesity or by other mechanisms possibly related to their disease and treatment.

Interpretation and generalisability

These results show that for research purposes in childhood cancer survivors the 3D method has advantages over 2D and can be validly used, but with caution. Provided the same equipment is used, it is possible to perform ultrasounds in more than one centre, still allowing for evaluation by one observer. In addition, the ultrasonographers do not have to interpret the ultrasound data when performing the measurements and the investigator analysing the 3D data can assess AFCs while having no knowledge about the medical history. This will minimize observer bias.
For counselling individual childhood cancer survivors, caution should be taken when interpreting antral follicle counts measured in 3D mode as being representative of ovarian reserve. Weighted Cohen's Kappa values showed that the agreement between 2D and 3D was low to moderate and it is therefore plausible that individual patients can be wrongly classified into a category with better or worse prognosis. Additional studies are required to demonstrate whether antral follicle counts measured in 3D mode agree well with the gold standard, i.e. histopathology of the ovary. For now, we recommend a full panel of ovarian reserve tests, including AMH, to advise patients on ovarian reserve.

9

Comparison of ovarian function markers in users of hormonal contraceptives during the hormone-free interval and subsequent natural early follicular phases

M.H. van den Berg
E. van Dulmen-den Broeder
A. Overbeek
J.W.R. Twisk
R. Schats
F.E. van Leeuwen
G.J. Kaspers
C.B. Lambalk

Hum Reprod.
2010 Jun;25(6):1520-7.

ABSTRACT

Objective

The aim of this study was to evaluate whether values of FSH, LH, estradiol, anti-Müllerian hormone (AMH), inhibin B, antral follicle count (AFC) and ovarian volume (OV) determined on day 7 of the hormone-free interval are similar to values measured on days 2–5 of two subsequent natural menstrual cycles. In addition, values measured on day 7 of the hormone-free interval were examined for the purpose of predicting values measured on days 2–5 during the second natural cycle.

Methods

In this study, 25 women using hormonal contraception provided a blood sample and underwent transvaginal ultrasound measurements on day 7 of the hormone-free interval and on cycle days 2–5 of two subsequent natural cycles. Changes were compared by repeated measures ANOVA and multivariate linear regression was used for prediction purposes.

Results

Mean (SD) age of the participants was 26.3 (6.2) years. Overall significant decreases in FSH and inhibin B and significant increases in AMH, AFC and ovarian volume values were measured after discontinuation of hormonal contraception ($P < 0.001$, $P = 0.04$, $P = 0.01$, $P < 0.001$ and $P = 0.004$, respectively). Significant changes occurred both from day 7 of the hormone-free interval to natural cycle 1 as well as from natural cycle 1 to natural cycle 2. FSH, AMH and AFC values measured during days 2–5 of natural cycle 2 could be predicted by the corresponding values measured on day 7 of the hormone-free interval.

Conclusion

Hormonal and ultrasound markers of ovarian function in hormonal contraception users measured at the end of the hormone-free interval do not seem to represent subsequent natural early follicular phase values. However, these values can, in some cases (FSH, AMH and AFC), be used to predict early follicular phase values using calculated prediction equations, which need to be validated in future research.

INTRODUCTION

A fully functional hypothalamic-pituitary-ovarian axis is one of the prerequisites for an adequate female reproductive function. Hypothalamic-pituitary-ovarian activity is traditionally assessed by reproductive hormones (FSH, LH and estradiol (E2)), and transvaginal ultrasound measurements (antral follicle count (AFC) and ovarian volume (OV)) during the early follicular phase of a natural menstrual cycle in women not using hormonal contraceptives. In addition, new hormonal markers, such as anti-Müllerian hormone (AMH) and inhibin B, are emerging and are increasingly used, but mainly in research settings. However, the majority of women of reproductive age use hormonal contraceptives [278] and it would be of practical value to be able to accurately assess reproductive function under such conditions.

The contraceptive effect of hormonal contraception is predominantly established by suppression of gonadotrophin secretion by the pituitary, which in turn inhibits ovarian activity resulting in arrested follicle growth and reduced hormone production [279 280]. The standard regimen of oral contraceptives includes 21 days of estrogen/progestin pills followed by a 7-day hormone-free interval in which a withdrawal bleeding is induced. During this interval recovery of pituitary-ovarian activity occurs as the pituitary gland begins to secrete gonadotrophins [281 282] resulting in follicular development and hormone production [283-285]. Indeed, it has been demonstrated that FSH, LH, E2 and inhibin B values increase during the hormone-free interval [286] and that follicle growth is established, even to dominance in some cases [283 285 287]. Although the events during the hormone-free interval have shown to resemble those during the early follicular phase of a natural cycle [287 288], it remains unclear whether hormone and ultrasound values on day 7 of the hormone-free interval resemble natural early follicular phase values.

A few studies have shown that FSH levels at the end of the hormone-free interval are similar to early follicular phase levels, although differences in maximum levels have been observed between different types of oral contraceptives [281 287 289 290]. Similarly, LH levels were also found to be comparable [281 289]. However, E2 levels were found to be significantly lower at the end of the hormone-free interval compared with early follicular phase values but were comparable with midfollicular phase values at the day of dominant follicle selection [281 290]. Values of AMH, AFC and ovarian volume were unfortunately not included in the aforementioned studies and results were not obtained from measurements performed prospectively in the same group of women. This may be a substantial limitation of these studies since the large inter-individual variability of the reproductive markers and the relatively small study groups may have made it difficult to adequately determine whether levels of reproductive markers at the end of the hormone-free interval are comparable to those measured during the early follicular phase.

To our knowledge only one study has compared reproductive markers in the hormone-free interval and in the early follicular phase within one group of women. These women, who served as a control group, were first measured during the early follicular phase of a natural menstrual cycle and subsequently started using oral contraception [291]. Data showed that FSH, LH, E2, AFC and ovarian volume values measured on days 3–5 of the natural cycle decreased significantly after six cycles

of oral contraceptives and were not comparable to values measured during the hormone-free interval. AMH values, however, remained unchanged. This implies that, except for AMH, values of reproductive markers obtained during the hormone-free interval do not resemble early follicular phase values.
However, no data are available on short-term changes in values of reproductive markers in women using hormonal contraceptives who subsequently discontinue this usage. This information is relevant to clinicians who counsel young women of reproductive age who are in need of information about their reproductive function, sometimes without an immediate wish to have children. It is known that a large proportion of these women use oral contraception as this is the most widely used method of contraception among women of reproductive age [278 292]. Therefore, from a clinical point of view it would be of interest to obtain a decisive answer as to whether women, who use oral contraceptives and who would like their reproductive function to be evaluated without having an immediate wish to have children, should stop taking these contraceptives for a period of time in order to adequately assess their reproductive function.
Therefore, we designed the current exploratory study to evaluate whether values of FSH, LH, E2, AMH, inhibin B, AFC and total ovarian volume determined on day 7 of the hormone-free interval are comparable to values measured on days 2–5 of two subsequent natural menstrual cycles. In addition, if these values were not similar, we investigated whether values measured on days 2–5 of the second natural cycle could be predicted from the corresponding values measured on day 7 of the hormone-free interval.

METHODS

Subjects

Study participants were recruited through advertisements on blackboards and in hallways of the VU University Medical Center, the VU University and several family practitioners in Amsterdam. To those interested in participating in the study, study information was given orally and in writing.
Inclusion criteria were female gender, age 18–40 years at study entry, hormonal contraceptive use for at least 3 months (either using the standard 21- /7-day regimen or an extended regimen) prior to the start of the study, an initial contraceptive indication for use of hormonal contraceptives, willingness to discontinue hormonal contraceptives use for at least two natural menstrual cycles and willingness to use methods of contraception other than hormonal contraception during these cycles. Exclusion criteria were virginity, history of endocrine disease such as thyroid dysfunction and history of ovarian or cranial surgery. During the study period, condoms were provided free of charge as an alternative method of contraception. All participants gave written informed consent. Furthermore, they were compensated financially for their participation.

Study design

The study was designed as a longitudinal prospective study and was approved by the Medical Ethics Committee of the VU University Medical Center in Amsterdam. Data were collected by a one time only questionnaire and by blood sampling and transvaginal ultrasound measurements of the reproductive organs on three occasions. Subjects who agreed to participate were first sent the questionnaire. After having received the filled-out questionnaire, an appointment was made for the first measurement. Since for the purpose of this study this measurement had to take place on the day of the hormone-free interval on which the influence of the hormonal contraception would be the least, this measurement was performed on the last day of the hormone-free interval, i.e. day 7. The second and third hospital visit were planned on days 2, 3, 4 or 5 of the two subsequent natural menstrual cycles.
If the first natural menstrual bleeding did not occur within 3 months of the first day of the withdrawal bleeding (following the discontinuation of hormonal contraception), participants were contacted by the researcher and referred to a gynecologist when desired. These amenorrheic participants were excluded from the study.
Participants were asked to monitor their basal body temperature (BBT) during the first natural menstrual cycle following discontinuation of hormonal contraception. An experienced gynecologist (C.B.L.) identified ovulation based on an obvious biphasic shift of around 0.3°C in the BBT. If ovulation could not be verified by the BBT chart, an elevated mid-luteal phase progesterone level (≥ 10 nmol/l) was confirmative for ovulation. Cycle length of the first completely evaluable natural cycle (i.e. the second natural cycle) was determined by counting the number of days from the first day of bleeding until the day before the next bleeding period.

Data collection

Questionnaire

The questionnaire was an adaptation of a well-tested questionnaire used by the Department of Epidemiology of the Netherlands Cancer Institute in a large-scale Dutch cohort study on long-term effects of ovarian stimulation for in vitro fertilization [293]. It addressed the following issues relevant to our study: socio-demographic characteristics, menstrual history and type and duration of current and past usage of hormonal contraceptives.

Hormonal assays

Blood samples were centrifuged for 10 min at 4°C (3000 rpm) within 30 min after venipuncture and frozen (–20°C) for storage until assayed. Laboratory screening was performed by the endocrine laboratory of the VU University Medical Center. All samples of one individual were analysed in the same run for each hormone. Plasma FSH levels were analysed by an immunometric assay (Delfia, Wallac, Turku, Finland), with a lower detection limit of 0.5 IU/l. The intra- and inter-assay coefficient of variation (CV) was 5 and 7%, respectively, at a concentration of 2 IU/l and 3 and 6%, respectively, at a concentration of more than 4 IU/l. Plasma LH levels

were determined by an immunometric assay (Delfia, Wallac, Turku, Finland), with a lower detection limit of 0.3 IU/l. The intra- and inter-assay CVs were 3 and 7%, respectively. E2 was measured by radioimmunoassay (Daisorin, Sallugia, Italy) with a lower limit of quantification of 18 pmol/l and an intra- and inter-assay CV of 5 and 10%, respectively. An ultra-sensitive immuno-enzymometric assay kit (Diagnostic Systems Laboratories, Webster, TX, USA) was used to measure the AMH in duplicate [391]. The limit of quantification was 0.1 µg/l. Intra- and inter-assay CV was 5 and 8%, respectively. Inhibin B was measured in duplicate by ultra-sensitive two-site enzyme immunoassays (Serotec, Oxford, UK). The lower limit of quantification was 15 pg/ml and the intra- and inter-assay CV was 5 and 9%, respectively.

Ultrasound

All transvaginal ultrasound measurements were performed by a specifically trained investigator (MvdB) using a 6 MHz transvaginal probe (EnVisor HD, Philips Medical Systems, Eindhoven, The Netherlands). An AFC was performed, counting the number of follicles sized 2–10 mm in both the right and left ovary. Furthermore, the volume of the left and right ovary was estimated from its length (L) and width (W) using the formula $(\pi LW^2)/6$, which assumes the ovaries to have a prolate ellipsoid shape. Ovarian volume was defined as the mean value of the right and left ovary.

Statistical analysis

Data were analysed using SPSS for Windows-version 14.0 (SPSS Inc., Chicago, IL, USA). Data, which were not normally distributed, were log-transformed prior to analysis and normality was checked again after this transformation. Changes in hormone levels (FSH, LH, E2, AMH, inhibin B) and ultrasound measurements (AFC and OV) were analysed by repeated measures ANOVA. When the overall change over time was significant, 'within subjects contrasts' analyses were used to detect which changes between two time points reached statistical significance. Furthermore, Pearson correlation coefficients were used to evaluate whether changes in hormone and ultrasound values were dependent on age, ethinyl estradiol dose of hormonal contraception or BMI. In order to assess whether hormone levels during the hormone-free period can act as markers for the number of ovarian follicles, we analysed whether AMH and AFC were correlated in conditions of hormonal contraceptive use as well as in conditions of natural cycles. Since AMH is currently considered to be the most promising marker of ovarian function [294 295], Pearson correlations with AFC were calculated for this variable only. In addition, by calculating Pearson correlations between levels of AMH at all three time points, the inter-cycle stability of AMH was evaluated. Finally, multivariable linear regression was used to predict hormone or ultrasound values in natural cycles from the corresponding values on day 7 of the hormone-free interval. In addition, it was evaluated whether the interaction between the age and the variable to be evaluated significantly contributed to the regression model. A P-value < 0.05 was considered to be statistically significant.

For the sample size calculation, the AFC variable was used. The sample size required to detect at least a change of four antral follicles [291] was calculated. The power to detect this change was set at 90% and the significance level at 0.05. Based on

the formula of Twisk et al. [218], it was estimated that a minimum of nine female participants would be needed to detect these changes in the number of follicles. Since other outcomes are also included in this study and to compensate for an expected dropout rate of 15% and an 'exclusion rate' after the first measurement of 10% (because of non-occurrence of first natural bleeding), we planned to enrol at least 30 women in the study.

RESULTS

Thirty women were recruited for the study and five dropped-out: one because of repeatedly not showing up, three as a result of polycystic ovarian syndrome induced amenorrhea, and one due to the presence of a dermoid cyst in the ovary which made it not possible to perform an AFC. As a result, the study population consisted of 25 women who all appeared to have ovulatory cycles (23 established by BBT curves and 2 by additional mid-luteal progesterone assessment). The characteristics of the participants are summarized in Table 1.

Table 1 Socio-demographic characteristics of the 25 study participants*

Age, years; mean (SD)	26.3 (6.2)
Body Mass Index, kg/m^2; mean (SD)	22.0 (1.8)
Type of hormonal contraception last used	
20 µg Estrogen monophasic pill	3 (12)
30-35 µg Estrogen monophasic pill	17 (68)
50 µg Estrogen monophasic pill	1 (4)
30-40 µg Estrogen triphasic pill	2 (8)
Other #	2 (8)
Duration of hormonal contraception last used, months; median (IQR †)	20 (57)
Regimen of last used hormonal contraception	
Standard 21-day/7-day regimen	21 (84)
Extended regimen	4 (16)
Cycle length, days; mean (SD)	28.8 (4.8)

** Values are the number (%) of women, unless indicated otherwise. # Ortho-Evra Patch (n = 1) and NuvaRing (n = 1). † IQR = interquartile range.*

The mean (SD) hormone and ultrasound values on day 7 of the hormone-free interval and on days 2, 3, 4 or 5 of the two subsequent natural menstrual cycles are shown in Table 2.

Table 2 Mean (SD) values of hormone levels and ultrasound characteristics on 3 chronological time points and significance of change scores between different time points[#]

	Day 7 HFI [†]	NC1 [†]	NC2 [†]	P-value NC1-HFI	P-value NC2-NC1	P-value NC2-HFI
FSH (U/l)	6.9 (3.0)	5.6 (2.0)	5.4 (2.1)	0.003*	0.18	< 0.001*
LH (U/l)	3.5 (1.7)	4.1 (1.6)	3.9 (1.5)	0.23	0.74	0.28
E2 (pmol/l)	90.4 (41.2)	95.4 (38.0)	106.6 (45.6)	0.63	0.27	0.30
AMH (µg/l)	2.0 (1.2)	2.0 (1.3)	2.6 (1.3)	0.91	0.01*	0.001*
Inhibin B (ng/l)	88.3 (47.5)	61.9 (32.1)	68.5 (36.8)	0.002*	0.62	0.09
Antral follicle count (no.)	19.9 (8.1)	22.0 (7.8)	23.9 (7.7)	0.02*	0.01*	< 0.001*
Ovarian volume (cm^3)	3.0 (1.2)	3.2 (1.3)	4.0 (1.4)	0.35	0.01*	0.01*

*[#] Analysed with repeated measures ANOVA. [†] HFI = hormone-free interval; NC1 = natural cycle 1; NC2 = natural cycle 2. * P < 0.05.*

Repeated measures ANOVA showed an overall significant decrease in FSH and inhibin B values and a significant increase in AMH, AFC and OV values after discontinuation of the pill (P < 0.001, P = 0.04, P = 0.005, P < 0.001 and P = 0.004, respectively). The overall effect of time regarding the other hormone values (LH and E2) appeared non-significant. Additional within subjects contrast analysis indicated that significant changes were measured from day 7 of the hormone-free interval to natural cycle 1 but also from natural cycle 1 to natural cycle 2 and from day 7 of the hormone-free interval to natural cycle 2 (Table 2). None of the changes appeared to significantly correlate with age, ethinyl estradiol dose or BMI (data not shown). Correlations between AMH and AFC values measured at day 7 of the hormone-free interval, at the first natural cycle and the second natural cycle all appeared to be significant (r = 0.50 (P = 0.02); r = 0.47 (P = 0.02); r = 0.67 (P = 0.001), respectively). Furthermore, the correlation between AMH at day 7 of the hormone-free interval and AMH at the first natural cycle was 0.75 (P < 0.001), whereas the correlations between AMH at the first and the second natural cycle and between AMH at day 7 of the hormone-free interval and AMH at the second natural cycle were 0.80 (P < 0.001) and 0.85 (P < 0.001), respectively.

Results of linear regression analysis (Table 3) revealed that FSH, AMH and AFC values measured on day 7 of the hormone-free interval significantly contributed to the prediction of values measured during the early follicular phase of natural cycle 2, all explaining about 72% of the variance.

Table 3 Results of regression analysis.

Dependent variable	Independent variable	Regression coefficient (95% CI)[#]	P-value	R^2
FSH at NC2 [†]	FSH at day 7 HFI [†]	0.589 (0.420, 0.758)	< 0.001*	0.71
AMH at NC2	AMH at day 7 HFI	0.945 (0.680, 1.210)	< 0.001*	0.72
Inhibin B at NC2	Inhibin B at day 7 HFI	-0.004 (-0.347, 0.340)	0.98	< 0.001
AFC at NC2	AFC at day 7 HFI	0.778 (0.543, 1.014)	< 0.001*	0.72
OV at NC2	OV at day 7 HFI	0.349 (-0.167, 0.866)	0.17	0.09

[#] Crude regression coefficients are reported. [†] HFI = hormone-free interval; NC2 = natural cycle 2.

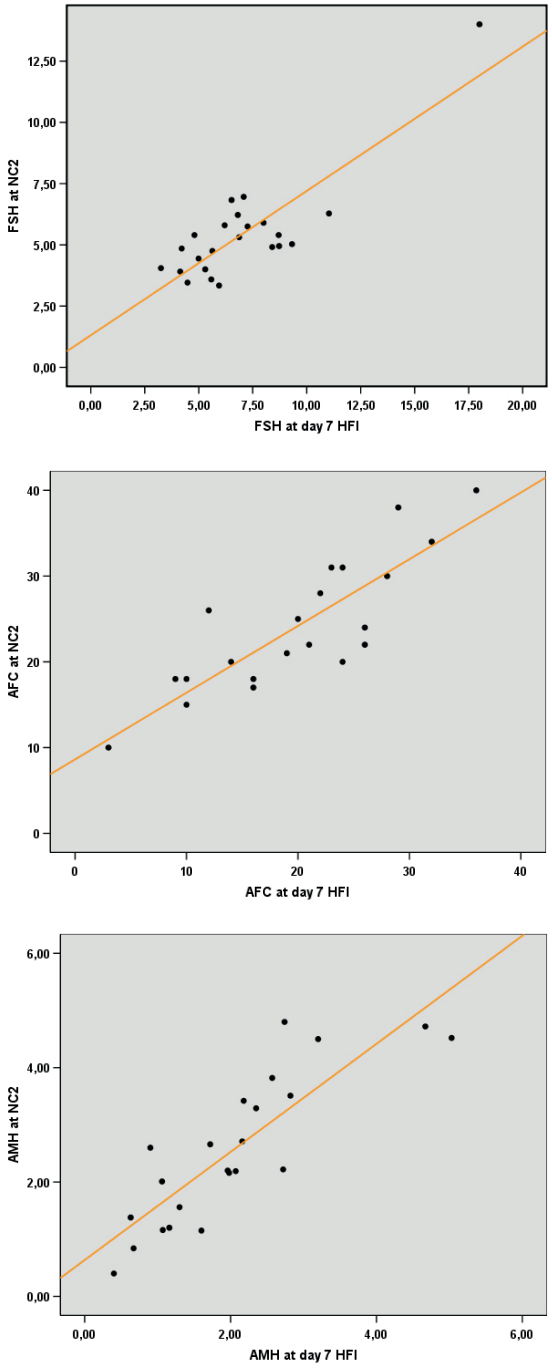


Figure 1 Scatter plot of FSH, AMH and AFC values measured on day 7 of the hormone-free interval (HFI) versus those measured during the second natural menstrual cycle (NC2).

516278-L-bw-Overbeek
Processed on: 19-1-2018

For FSH it was found that adding age and the interaction between age and FSH at day 7 to the model revealed that the interaction with age was significant (P = 0.02), increasing the explained variance to 82%. This analysis led to the following regression equation: FSH at NC2 = 5.922 – (0.347 × FSH at day 7 HFI) – (0.112 × age) + (0.026 × FSH at day 7 HFI × age), where HFI is the hormone-free interval and NC2, the natural cycle 2.
For example, when FSH values of a 30-year old woman measured on day 7 of the hormone-free interval is 7.0 U/l, the expected value during the early follicular phase of the second natural menstrual cycle is 5.922 – (0.347 × 5.0) – (0.112 × 30) + (0.026 × 5. × 30) = 5.6 U/l. Calculated regression equations for AMH and AFC were: [AMH at NC2 = 0.636 + (0.945 × AMH at day 7 HFI)] and AFC (AFC at NC2 = 8.629 + (0.778 × AFC at day 7 HFI), with HFI = hormone-free interval and NC2 = natural cycle 2). Neither inhibin B nor OV on day 7 of the hormone-free interval could predict the corresponding values during the early follicular phase of natural cycle 2.
Correlations between values of FSH, AMH and AFC measured on day 7 of the hormone-free interval and those obtained during the second natural menstrual cycle were high (see R^2 values in Table 3). These correlations are illustrated in the graphs in Figure 1. The correlation coefficient stated in the FSH-graph remained significant when the elevated FSH extreme (> 12 U/l) was excluded (R = 0.53, P = 0.01).

DISCUSSION

This study demonstrates that most hormonal and ultrasound markers of ovarian function measured at the end of the hormone-free interval in women using hormonal contraception do not seem to represent natural early follicular phase values. FSH and inhibin B values appear to decrease significantly when contraception use is discontinued, whereas values of AMH, AFC and OV increase and LH and E2 did not change significantly. These changes were not dependent on age, ethinyl estradiol dose or BMI. Furthermore, regression models allow FSH, AMH and AFC values measured on day 7 of the hormone-free interval to predict the corresponding values during the early follicular phase of the second natural cycle.
The results of this study may provide clinicians and researchers knowledge on whether ovarian function can be assessed in women who are using hormonal contraception and in need of information on their fertility potential without having an immediate wish to have children. This for example could be the case for young women of reproductive age who have previously been treated for cancer. It is known that these women are in need of information about their reproductive function as most of them are aware that former treatment regimes might have caused damage to their reproductive system [296-298]. Since a large proportion of these females use oral contraceptives it could be decided to measure reproductive markers on day 7 of the hormone-free interval. However, the results of our study show that in that case caution should be taken when interpreting values of these markers measured at this point in time. Moreover, future research in which female cancer survivors are included is needed in order to establish the exact value of reproductive markers measured on day 7 of the hormone-free interval for this group of women.

The increased level of FSH in this study on day 7 of the hormone-free interval implies a limited negative feedback during this interval. This is most likely to occur during the first few days of this interval since high FSH levels coincide with elevated inhibin B levels on day 7 compared with early follicular phase values. Most probably, this increased level of inhibin B on day 7 is a reflection of a pronounced ovarian response to the earlier generated increase in FSH, although at this stage multiple follicle growth could not be confirmed by our data. However, it is known that multiple pregnancy can occur as a consequence of multiple follicle growth shortly after interrupting hormonal contraception [299 300].

Our results differ to some extent with the published data. In contrast to our study, previous studies report maximum levels of FSH at the end of the hormone-free interval to be similar to early follicular phase levels in the natural cycle, although maximum levels differ between different types of contraceptive pills due to a varying ethinyl estradiol component [281 287 289 290]. Furthermore, previous studies found levels of E2 to be significantly lower at the end of the hormone-free interval compared with early follicular phase values [281 290], whereas LH values were similar [281 289]. We found no changes in both E2 and LH and a higher value of FSH at the end of the hormone-free interval. This may be explained by methodological differences between the previous studies in which between-group comparisons were used, whereas we measured changes within subjects.

To our knowledge, only one study has compared tests for measuring ovarian responsiveness within a group of subjects [291]. In this study these tests were performed among 15 healthy women before and after 6 months of oral contraceptive use. We found comparable results with respect to AFC and OVs, i.e. lower values during the hormone-free interval compared with early follicular phase values. However, with respect to hormone values contradicting results were found. In the study by Somunkiran et al., no changes in the AMH values were found, while we found AMH to be significantly lower during the hormone-free interval. Furthermore, we found FSH values to be significantly higher during this interval and LH and E2 values to be unchanged while they found FSH, LH and E2 to be significantly lower during the hormone-free interval. Apparently, it either seems to make a difference whether measurements during the early follicular phase were performed prior to or following measurements during the hormone-free interval or whether the measurements in the hormone-free interval were performed after a relatively short (6 months in the study of Somunkiran et al.) or a prolonged period of contraceptive use (20 months on average in our study). It should, however, be noted that the results in the study of Somunkiran et al. are based on a relatively small sample size [291].

In addition, a possible explanation for the contradicting results regarding FSH, LH and E2 may be found in the timing of the measurement during the hormone-free interval. It has been demonstrated that during the 7-day hormone-free interval FSH, LH and E2 values increase significantly. However, the start of this increase varies between different hormones and the type of oral contraceptives used [285 288 301]. It has been shown that the suppressive effect of oral contraceptives is prolonged with higher doses of ethinyl estradiol. Despite this, levels of reproductive hormones on day 7 of the hormone-free interval appear to be higher with increasing doses of ethinyl estradiol [281 285]. All women in the study of Somunkiran et al. [291] used 35

µg E2 pills while in our study the majority of the women used either 30 µg E2 pills (n = 17) or 20 µg E2 pills (n = 3). Therefore, the rise in hormone levels during the hormone-free interval is expected to occur later in the Somunkiran study compared with our study. Moreover, measurements were performed on day 5 of the hormone-free interval in the study by Somunkiran while this was done on day 7 in our study. This might explain why significantly lower FSH values during the hormone-free interval compared with natural cycle values were found in the study by Somunkiran et al., while we found the opposite.

Recent studies indicate that serum levels of AMH are relatively constant throughout the menstrual cycle or that at least large physiological fluctuations are not seen [21 22 239 302]. This is likely to be a reflection of the rather constant size of smaller FSH-sensitive antral follicles that typically produce AMH [16]. Development of these follicles is not yet under full control of FSH but when they enter the stage of FSH-responsiveness AMH production drops [303 304]. Furthermore, several studies show a drop in the AMH levels concurrent with stimulation with exogenous FSH for example in women undergoing IVF treatment [305 306]. This drop in AMH is currently interpreted as an indication that the number of smaller pre-antral follicles temporarily decreases. Therefore, the most plausible explanation for our observation that AMH increases after discontinuation of hormonal contraception is that the cohort of smaller antral follicles increases monthly and that prolonged suppression of FSH by means of hormonal contraception to some extent prevents pre-antral and small antral follicle formation, i.e. the class of follicles that under normal circumstances typically contributes to serum AMH levels. So, somehow previous use of hormonal contraceptives influences AMH values. However, even in our study inter-cycle AMH correlations remained extremely high indicating that AMH seems to be a reliable marker for measuring ovarian function, a statement supported by other studies as well [294 295 307]. Moreover, the significant correlations between AMH and AFC found in our study as well as in other studies [308] seem to endorse this statement.

In our study it was shown that, except for E2 and LH, all markers of reproductive function changed significantly after discontinuation of hormonal contraception. One can, however, argue that the individual differences between values measured during the hormone-free interval and the natural cycle are not relevant from a clinical point of view since changes occurred within normal reference values in most cases. However, future research involving more and different types of subjects (e.g. women with proven sub- or infertility, or older women) should further examine this clinical relevance. It should for example be investigated to what extent women would be diagnosed differently regarding their fertility status when measurements on day 7 of the hormone-free interval are compared with measurements during natural cycles. It might be the case that measuring these women on either one of these time points would not bring on any different clinical consequences. However, from a research point of view, differences measured at the two time points indeed might be of relevance, since one has to be cautious in interpreting mean values of study groups, especially if these groups include both hormonal contraception users and non-users. If values of hormonal contraception users and non-users are combined and reported as representative of a study group, this value could represent an

over- or underestimation of the true value, depending on the number of hormonal contraception users in the study group concerned. Therefore, discontinuation of hormonal contraception use seems justified in research involving women on hormonal contraception, particularly when comparing mean values of two or more study groups.

Thus, results of our study indicate that at this time, to assess ovarian function by a full panel of reproductive markers, it seems most appropriate to discontinue oral contraceptive use for at least two months. However, this study does provide regression equations from which values of FSH, AFC and AMH during the natural menstrual cycle can possibly be predicted from values measured during the hormone free interval. These equations should however be validated first in future research in larger groups of women (including women with apparently normal ovarian function as well as women with a decreased ovarian function) before they can be used for clinical or research purposes.

There are some limitations to the present study. First of all, different types of hormonal contraception were used. This might have influenced our results since maximum levels of reproductive hormones measured on day 7 of the hormone-free interval seem to depend on the type of contraceptives used [285 288 301]. Previous studies have shown that higher doses of ethinyl estradiol lead to prolonged periods of hormonal suppression but also to significantly higher levels of FSH on day 7 of the hormone-free interval [281 285]. Unfortunately, due to the small sample size in our study we were not able to reliably evaluate the effect of the different types of hormonal contraception used. Future research should investigate whether oral contraceptive pills with different doses of ethinyl estradiol bring on different changes in reproductive markers after discontinuation of these pills.

Secondly, the values of the reproductive markers observed in this study may have been subject to inter-cycle variability. Indeed, it has been demonstrated that inter-cycle variability is present both in hormonal as well as ultrasound markers of ovarian function [305 308-313]. However, the statistical analyses used in this study take into account this type of variability. Since we were able to detect significant changes despite the acknowledged inter-cycle variability, it can be stated that the significant differences found in this study are likely to reflect true biological changes.

Thirdly, in this study significant changes in reproductive markers did not only occur between the first two time points, i.e. day 7 of the hormone-free interval and first natural cycle, but for some markers values continued to change significantly between the first and second natural menstrual cycle. Therefore, it could be that the values of reproductive markers of each new cohort of follicles might continue to increase, decrease or revert to baseline even after the second natural menstrual cycle. Future research in which longer follow-up protocols are used is therefore recommended and is currently under way. Moreover, data obtained from such research can be used to validate the regression equations provided by this study, enabling clinicians and researchers to reliably predict early natural cycle values from values measured during the hormone-free interval. However, it should be noted that the results of our study do not contribute to the question whether or not ovarian function or the occurrence of primary ovarian insufficiency can be accurately predicted from values of reproductive markers measured on day 7 of the hormone-free interval. We merely

studied whether values of reproductive markers were similar during the pill-free period and natural menstrual cycles. Even when these markers are measured during the early follicular phase (as this is clinical practice in normal circumstances), it is still a matter of debate whether these markers can indeed reliably be used for these kind of prediction purposes [3 314].

Finally, in our study all participants were ostensibly healthy and appeared to have normal ovarian function values. Only one (older-aged) woman had elevated FSH concentrations with corresponding lower AFC values when compared with the other participating women. It is of importance to investigate in future studies if the results we found in this study also apply to women with primary ovarian insufficiency.

In conclusion, this study has demonstrated that hormonal and ultrasound markers of ovarian function in hormonal contraception users measured at the end of the hormone-free interval do not seem to represent subsequent natural early follicular phase values. These results are of importance when measuring ovarian function of women who use hormonal contraception as they indicate that, when reproductive markers are measured at day 7 of the hormone-free interval, caution should be taken when interpreting the values of these markers measured at this point in time.

516278-L-bw-Overbeek
Processed on: 19-1-2018

10

Long-term effects of childhood cancer treatment on ovarian function markers: results of the DCOG LATER-VEVO study

A. Overbeek
M.H. van den Berg
C.B. Lambalk
G.J.L. Kaspers
D. Bresters
M.M. van den Heuvel-Eibrink
L.C. Kremer
J.J. Loonen
H.J.H. van der Pal
W.J.E. Tissing
C.M. Ronckers
A.B. Versluys
M. van der Heiden-van der Loo
A.C. Heijboer
M. Hauptmann
J.W.R. Twisk
J.S. Laven
C.C.M. Beerendonk
F.E. van Leeuwen
E. van Dulmen-den Broeder
on behalf of the DCOG
LATER-VEVO study group

Submitted

ABSTRACT

Background
Previous studies on ovarian function in female childhood cancer survivors (CCSs) had several drawbacks, including lack of an extensive panel of ovarian function tests, reliance on self-reported data, and lack of a (large) comparison group. We investigated the (dose-related) impact of individual chemotherapeutic (CT) agents and radiotherapy (RT) fields on various markers of ovarian function in a nationwide cohort of CCSs.

Method
The DCOG LATER-VEVO study includes adult female five-year CCSs from seven Dutch Pediatric Oncology/Haematology Centres, diagnosed 1963-2002 (n=1,106, median attained age 27.8 years) and a comparison group (n=819). Reproductive history was assessed by questionnaire; ovarian function by anti-Müllerian hormone (AMH), follicle-stimulating hormone (FSH), inhibin B levels, and antral follicle counts (AFC).

Findings
Overall, 7.0-17.7% of CCSs and 2.4-13.6% of controls had a reduced ovarian function (depending on marker used). Above age 35, the proportion of CCSs with abnormal ovarian function markers was significantly higher compared to controls for AMH (26% vs. 4%), AFC (20% vs. 3%), and inhibin B (42% vs. 16%). For AMH and FSH significant differences between both groups were also found below 35 years. Among CCSs treated with alkylating chemotherapy and gonadotoxic radiotherapy, aged 18-34 years, the proportions with low AMH, low AFC, high FSH or low inhibin B levels were 46%, 19%, 27%, and 43%, respectively, while these were at most 17% in those treated with alkylating CT only. Cyclophosphamide, procarbazine, a composite group of “other alkylating agents”, dactinomycin, doxorubicine, mitoxantrone, spinal RT, abdominal/pelvic RT, and total body irradiation were multivariably associated with ≥ 1 abnormal ovarian function marker. Dose-effect relationships were established for procarbazine and abdominal/pelvic RT.

Interpretation
Although subgroups of CCSs are at high risk of reduced ovarian function, it is reassuring that the majority is not. Our results are important for counselling CCSs and future patients regarding parenthood and fertility preservation.

INTRODUCTION

A compromised reproductive function is a frequently encountered late effect of treatment in the growing group of female childhood cancer survivors (CCSs), with major impact on quality of life [39 315]. Cancer treatment, in particular alkylating chemotherapy, procarbazine, TBI and abdominal-pelvic radiotherapy, may damage the hypothalamic-pituitary-ovarian-uterine axis, clinically leading to delayed or arrested puberty, infertility, subfertility and adverse pregnancy outcomes [39 43 45 57 64 127 316]. Importantly, treatment can reduce fertile life span and induce premature menopause, since it may deplete or accelerate the decline of the non-renewable pool of primordial follicles in the ovary (ovarian reserve) [17 44]. Knowing who is at risk is of a decreased ovarian reserve, and thus, a shorter than anticipated window of opportunity to conceive, is important to adequately counsel CCSs and future patients regarding family planning and fertility preservation.

Optimal assessment of ovarian oocyte count, mostly referred to as reserve, would be histological evaluation. This, however, cannot be done in both the clinical and research setting. Indirect, surrogate markers such as follicle stimulating hormone (FSH), inhibin B, anti-Müllerian hormone (AMH), and antral follicle count (AFC) are therefore used. Use of more than one marker appears useful in delineating specific stages in the natural transition to menopause [317]. Values of AMH and AFC have shown to decrease gradually with age, in contrast to FSH (the traditional marker of ovarian reserve), which does not start to rise until ovarian reserve is significantly reduced leaving a limited reproductive window if any. Currently, AMH and AFC are considered the most promising markers for assessing ovarian reserve, also in CCSs [318]. Studies have shown a highly significant, positive correlation between serum AMH and AFC and ovarian reserve assessed by manual counts of non-growing follicles in ovarian tissue samples [29].

Previous studies addressing reproductive late effects in female CCSs, however, often lacked a full panel of clinical ovarian function tests. In addition, they often relied on self-reported data only, had incomplete follow-up, or lacked a (large) comparison group [45 53 60 61 118 316].

The current study aimed to identify treatment-related factors associated with a reduced ovarian function in a large cohort of female long-term adult CCSs using, for the first time, an extensive set of clinical ovarian function markers. Moreover, associations between these markers were evaluated.

METHODS

Study design and patients

The DCOG LATER-VEVO study is a nationwide retrospective cohort study evaluating fertility, ovarian reserve and premature menopause among adult female 5-year CCSs, treated between 1963-2002 in the Netherlands, by questionnaire, blood sampling and a transvaginal ultrasound measurement of the ovaries. The study population consisted of CCSs from the DCOG-LATER cohort (n=2,237) and of a comparison group including sisters of participating CCSs as well as women from the general population

516278-L-bw-Overbeek
Processed on: 19-1-2018

(GP controls). Eligibility criteria and detailed study characteristics have been described previously and are summarized in Supplemental File 1 [271]. Approval was obtained from the relevant medical ethics committees and written informed consent from all participants.
Of the 1,749 eligible and invited CCSs and the 1,201 invited controls, 1,106 (63%) and 819 (68%), respectively, participated by completing the questionnaire (Figure S1). Of these participating CCSs and controls, 564 (51%) and 429 (52%), respectively, agreed to additionally provide a blood sample and/or have an ultrasound performed (clinical CCSs and clinical controls). However, for clinical assessments all women with a hormone-releasing intra-uterine device (n=52), previous ovarian surgery (n=35), and aged ≥ 52 years (n=19), were excluded. Therefore, ultimately, 552 clinical CCSs and 387 clinical controls were included in analyses. All provided a blood sample and 794 (455 CCSs, 339 controls) also underwent an ultrasound measurement.

Data collection procedures

The questionnaire addressed socio-demographic characteristics, life-style behaviour, virginity status, parity, menstrual cycle characteristics, menopause, and use of exogenous reproductive hormones.
Hormonal and ultrasound measurements were performed on cycle day 2-5 of a natural menstrual cycle or at any convenient moment in case of amenorrhea (no menses > six months). Women on hormonal contraceptives (HCs) were asked to stop HC use at least two months prior to study measurements. Those not discontinuing HCs were measured on day seven of the pill-free or ring-free week. Transvaginal ultrasound measurements were performed by trained personnel to assess the number of antral follicles sized 2-10 mm in both ovaries. Three-dimensional images of the ovaries were stored, enabling the actual AFC assessment to be performed by a single assessor retrospectively, using specialized software. Details of laboratory analyses and the ultrasound protocol are described in Supplemental Files 2 and 3, respectively.

Childhood cancer treatment

Details on prior cancer diagnosis and treatments, given for the initial malignancy, recurrences, and any known new primary malignancies until time of study evaluation, were collected from original medical files. Cumulative doses of each CT agent, and of RT administered to six body sites (cranium, spine, thorax, lower abdomen/pelvis, total body irradiation (TBI), other) were calculated for dose-effect analyses. Table S4.1 describes the exposure to specific CT agents among study participants and non-participants, whereas Table S4.2 describes percentile distributions of cumulative doses of specific CT agents and RT body sites.

Outcomes

Due to the absence of a gold standard marker, ovarian function was assessed by four surrogate markers: AMH, AFC, FSH, and inhibin B. Associations between these markers were evaluated using AMH as the ‘reference marker’ since it is currently

considered to be the most accurate and robust marker as fluctuations throughout the menstrual cycle appear small and, unlike AFC, it is not subject to inter-observer variation nor does it require transvaginal ultrasound measurement [319].

Statistical analyses

Differences between participating CCSs and controls were assessed using the Mann-Whitney-U and Pearson Chi-square test. Moreover, differences in characteristics between study participants, non-participants, deceased patients and those lost to follow-up, as well as between clinical and questionnaire-only CCSs and controls, were evaluated to assess the introduction of potential bias into the study (Tables S1.1 and S1.2).
Overall effect of treatment was assessed by comparing all CCSs with controls. After log-transformation of the four ovarian function markers, linear regression analyses were performed, one for each outcome. Effects of specific treatments were assessed using three different regression models: model 1 included four broad treatment subgroups based on their presumed gonadotoxicity (as reported in literature) [58 67 320], model 2a and 2b included accumulated scores for alkylating agent exposure (Alkylating Agent Dose (AAD) score [321] and Cyclophosphamide Equivalent Dose (CED) score) [109], and model 3 included (doses of) single CT agents and RT body sites. All model building procedures were based on linear regression analyses and details are summarized in Supplemental File 5a. To facilitate clinical interpretation of our data, logistic regression analyses were also performed, both for overall and specific effects of treatment, using the same variables as selected in the corresponding linear regression models. Specific cut-off levels were used to define low AMH, low AFC, low inhibin B, and high FSH levels. Since both AMH and AFC gradually decline with increasing age [258 322], age-specific cut-off values were calculated for these markers based on the included control population. Participants were categorised as having low AMH and low AFC values in case their value was more than 2SD below the mean of the corresponding age subgroup of controls (Supplemental File 5b). FSH, on the other hand, remains relatively stable until a steep rise in the last decade before menopause. Inhibin B follows this pattern [323]. Therefore, for FSH and inhibin B fixed cut-off values were used (>10 U/l and < 20 ng/l, respectively). All regression analyses were adjusted for age at study, time since diagnosis, pubertal status at treatment (i.e. menarche before/during/after treatment), smoking status, current body mass index, and HC use at clinical measurements. Analyses were performed using SPSS for Windows version 22.0 (SPSS, Chicago, IL).

RESULTS

Median age at study was 27.8 (interquartile range (IQR) 11.6) and 31.3 (12.4) years for participating CCSs (n=1,106) and controls (n=819), respectively (p<0.001) (Table 1).

Table 1 Characteristics of study participants*

	CCSs N= 1,106	Controls N=819	P value
Age at study, yrs; median (IQR)	27.8 (11.6)	31.3 (12.4)	< 0.001
≥ 18.0-24.9 yrs	400 (36.2)	213 (26.0)	
≥ 25.0-29.9 yrs	275 (24.9)	152 (18.6)	
≥ 30.0-34.9 yrs	178 (16.1)	170 (20.8)	
≥ 35.0-39.9 yrs	131 (11.8)	165 (20.2)	
≥ 40 yrs	122 (11.0)	118 (14.4)	
Education †			< 0.001
Low	97 (8.8)	26 (3.2)	
Medium	682 (62.2)	363 (44.8)	
High	318 (29.0)	422 (52.0)	
Current smoker			0.41
No	914 (82.6)	675 (82.4)	
Yes	168 (15.2)	138 (16.9)	
BMI, kg/m^2; median (IQR)	23.0 (5.6)	23.0 (4.9)	0.73
Use of exogenous reproductive hormones at time of study invitation			0.01
No	624 (56.4)	509 (62.1)	
Yes	482 (43.6)	310 (37.9)	
Primarily as contraception	413 (37.3)	303 (37.0)	
Primarily as HRT	63 (5.7)	6 (0.7)	
Primarily for other purposes	6 (0.5)	1 (0.1)	
Childhood cancer diagnosis			
Leukemias	386 (35.4)	--	
Lymphomas	176 (16.2)	--	
Renal tumours	124 (11.4)	--	
CNS tumours	113 (10.4)	--	
Soft tissue tumours	75 (6.9)	--	
Bone tumours	70 (6.4)	--	
Neuroblastomas	68 (6.2)	--	

Table 1 Characteristics of study participants* - Continued

	CCSs N= 1,106	Controls N=819	P value
Germ cell tumours	45 (4.1)	--	
Other malignant epithelial neoplasms	19 (1.7)	--	
Hepatic tumours	7 (0.6)	--	
Retinoblastoma	5 (0.5)	--	
Other and unspecified malignant neoplasms	1 (0.1)	--	
Age at diagnosis, yrs; median (IQR)	6.4 (8.4)	--	
Pubertal status at diagnosis			
Menarche > 3 years before diagnosis	720 (67.7)	--	
Menarche within 3 years of diagnosis	295 (27.7)	--	
Menarche > 3 years after diagnosis	49 (4.6)	--	
Time since diagnosis, yrs; median (IQR)	21.0 (11.7)	--	
Type of treatment based on presumed gonadotoxicity ¶			
No lower abdominal/pelvic RT and no alkylating CT	466 (42.2)	--	
Alkylating CT, but no lower abdominal/pelvic RT	517 (46.8)	--	
Lower abdominal/pelvic RT, but no alkylating CT	64 (5.8)	--	
Alkylating CT and lower abdominal/pelvic RT	58 (5.2)	--	
AAD-score			
0	526 (48.0)	--	
1	111 (10.1)	--	
2	185 (16.9)	--	
3	168 (15.3)	--	
≥ 4	106 (9.7)	--	
CED-score, mg/m^2			
0	530 (48.4)	--	
> 0 to < 4000	231 (21.1)	--	
≥ 4000 to 8000	154 (14.1)	--	
≥ 8000	179 (16.4)	--	

Table 1 Characteristics of study participants* - Continued

	CCSs N= 1,106	Controls N=819	P value
Ever had sexual intercourse			< 0.001
Yes	965 (87.1)	772 (94.3)	
No	141 (12.9)	47 (5.7)	
Age at study ≥ 18.0-24.9 yrs	81/400 (20.3)	27/213 (12.7)	
Age at study ≥ 25.0-29.9 yrs	25/275 (9.1)	6/152 (3.9)	
Age at study ≥ 30.0-34.9 yrs	14/178 (7.9)	5/170 (2.9)	
Age at study ≥ 35.0-39.9 yrs	13/131 (9.9)	6/165 (3.6)	
Age at study ≥ 40 years	8/122 (6.6)	3/118 (2.5)	
Given birth			< 0.001
No	736 (67.5)	457 (55.8)	
Yes	370 (33.5)	362 (44.2)	
Age at study ≥1 8.0-24.9 yrs	33/400 (8.3)	15/213 (7.0)	
Age at study ≥ 25.0-29.9 yrs	76/275 (27.6)	46/152 (30.3)	
Age at study ≥ 30.0-34.9 yrs	104/178 (58.4)	89/170 (52.4)	
Age at study ≥ 35.0-39.9 yrs	82/131 (62.6)	108/165 (65.5)	
Age at study ≥ 40 years	75/122 (61.5)	104/118 (88.1)	
Menopausal status			
Postmenopausal ǂ	82 (7.4)	22 (2.7)	< 0.001
Surgically-induced §	6 (0.5)	3 (0.4)	
Not surgically-induced (naturally occurring or CT-/RT-induced)	76 (6.9)	19 (2.3)	
Age at study ≥ 18.0-24.9 yrs	12/400 (3.0)	1/213 (0.5)	
Age at study ≥ 25.0-29.9 yrs	14/275 (5.1)	0/152	
Age at study ≥ 30.0-34.9 yrs	11/178 (6.2)	1/170 (0.6)	
Age at study ≥ 35.0-39.9 yrs	14/131 (10.7)	2/165 (1.2)	
Age at study ≥ 40 years	31/122 (25.4)	18/118 (15.3)	
With reported menopause before age 40 years	5/31 (16.1)	0/18 (0)	
Not surgically-induced premature menopause (overall rate)	56/1106 (5.1)	4/819 (0.5)	

CCSs: childhood cancer survivors; AAD: alkylating agent score; BMI: body mass index; CED: cyclophosphamide equivalent index; CNS: central nervous system; CT: chemotherapy; HRT: hormone replacement therapy; IQR: interquartile range; RT: radiotherapy. For some variables total N might not correspond with the number mentioned in the heading of the table because of missing data.

** Values represent the number (%) of women, unless indicated otherwise. IQR = Interquartile range.*

† Categorised as low: up to and including lower technical and vocational training; medium: up to and including secondary technical and vocational training; high: up to and including higher technical and vocational training and university.

[‡] *Postmenopausal women are defined as those women who have reported not to have had any menses in the past 12 months as well as those reported to use HRT at time of study invitation.*

[§] *Surgically-induced menopause is defined as cessation of menses caused by bilateral oophorectomy.*

[¶] *Reported treatment includes all treatments received up until time of study evaluation (i.e. treatment for initial diagnosis as well as for possible recurrences and subsequent cancers). For detailed information regarding the treatments received by the participating CCSs see Table S4.1.*

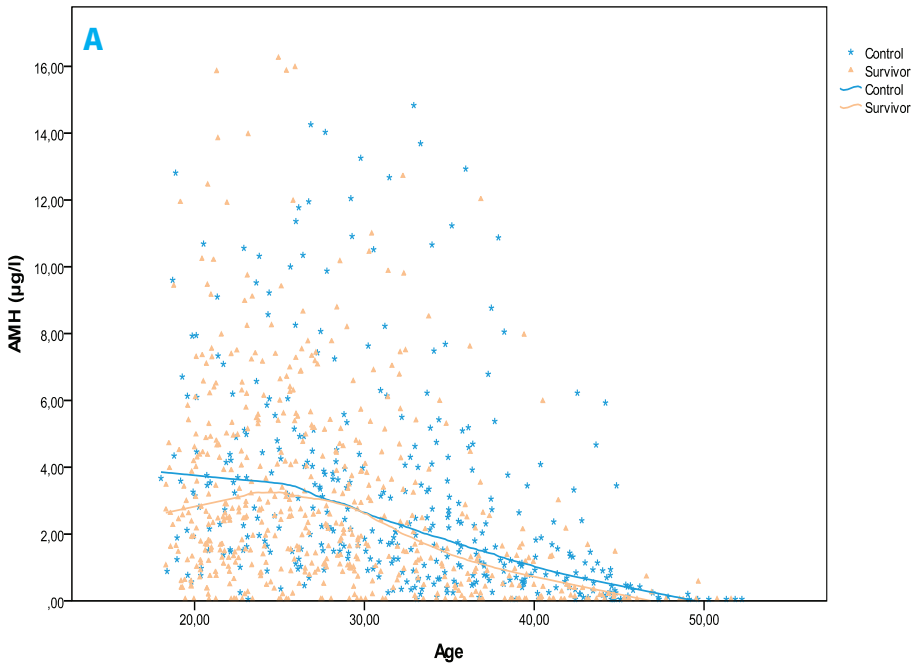
[‡] *The AAD-score was calculated by adding the tertile score (1, 2, or 3) for each of the alkylating agents given to a particular CCS [321]. An AAD-score of zero was assigned to non-exposed CCSs. The CED-score was calculated using a specific equation developed by Green et al. [109]*

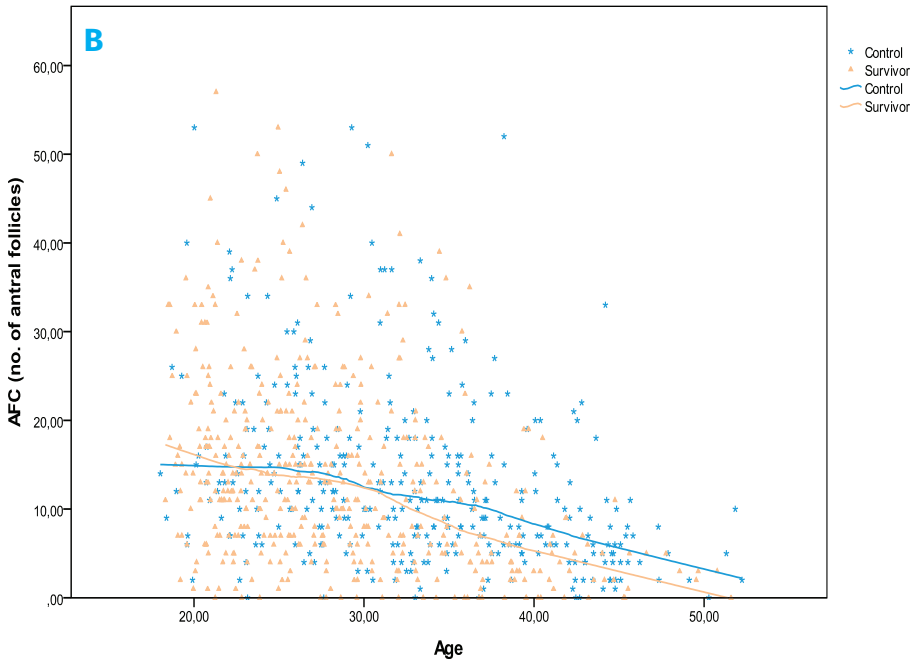
Compared to controls, CCSs were less likely to be highly educated (29.0 vs. 52.0%, p < 0.001) and more likely to use hormone replacement therapy. Median (IQR) age at diagnosis was 6.4 (4.8) years, and median time since diagnosis 21.0 (11.7) years. A large proportion of CCSs (42%) did not receive any presumed gonadotoxic CT and/or RT and almost half (48%) had both an AAD-score and a CED-score of zero.
CCSs were more likely to have had sexual intercourse (87.1 vs 94.3%, p<0.001), and to have given birth at time of study (33.5 vs 44.2%, p<0.001), the latter particularly in those > 40 years of age. In addition, more CCSs had reached menopause despite a younger average age at study and had a non-surgically-induced premature menopause (< 40 years; 5.1% and 0.5%, respectively (p<0.001). Menopause was evaluated as the composite outcome of reported amenorrhea and use of hormonal agents for menopausal symptoms.
Between participating and non-participating CCSs type of treatment differed significantly, but differences were clinically not relevant (Table S1.1). Clinical CCSs appeared significantly younger than questionnaire-only CCSs while, clinical controls were significantly older than questionnaire-only controls (Table S1.2). Moreover, birth rate was lower among both the clinical CCSs and controls compared to questionnaire-only CCSs and controls.

Overall effect of treatment

Median (IQR) AMH, AFC, FSH, and inhibin B levels were 2.1 (3.6) µg/l, 12.0 (13.0) follicles, 5.7 (2.9) U/l, and 68.0 (58.0) ng/l for CCSs and 1.9 (3.3) µg/l, 11.0 (11.0) follicles, 5.9 (2.7) U/l, and 73.0 (56.0) ng/l for controls, respectively. In both groups, AMH, AFC and inhibin B levels declined with increasing age, while FSH increased (Figure 1).

Overall, linear regression analysis showed that AMH, AFC, FSH and inhibin B levels did not differ significantly between CCSs and controls (data not shown). However, stratified by age, the proportion of CCSs with abnormal ovarian function markers was significantly higher compared to controls for all markers (except FSH) in the two highest age groups (i.e. > 35 years) (Table 2).


A
AMH (µg/l)
16,00
14,00
12,00
10,00
8,00
6,00
4,00
2,00
,00
20,00
30,00
40,00
50,00
Age
Control
Survivor
Control
Survivor


B
AFC (no. of antral follicles)
60,00
50,00
40,00
30,00
20,00
10,00
,00
20,00
30,00
40,00
50,00
Age
Control
Survivor
Control
Survivor

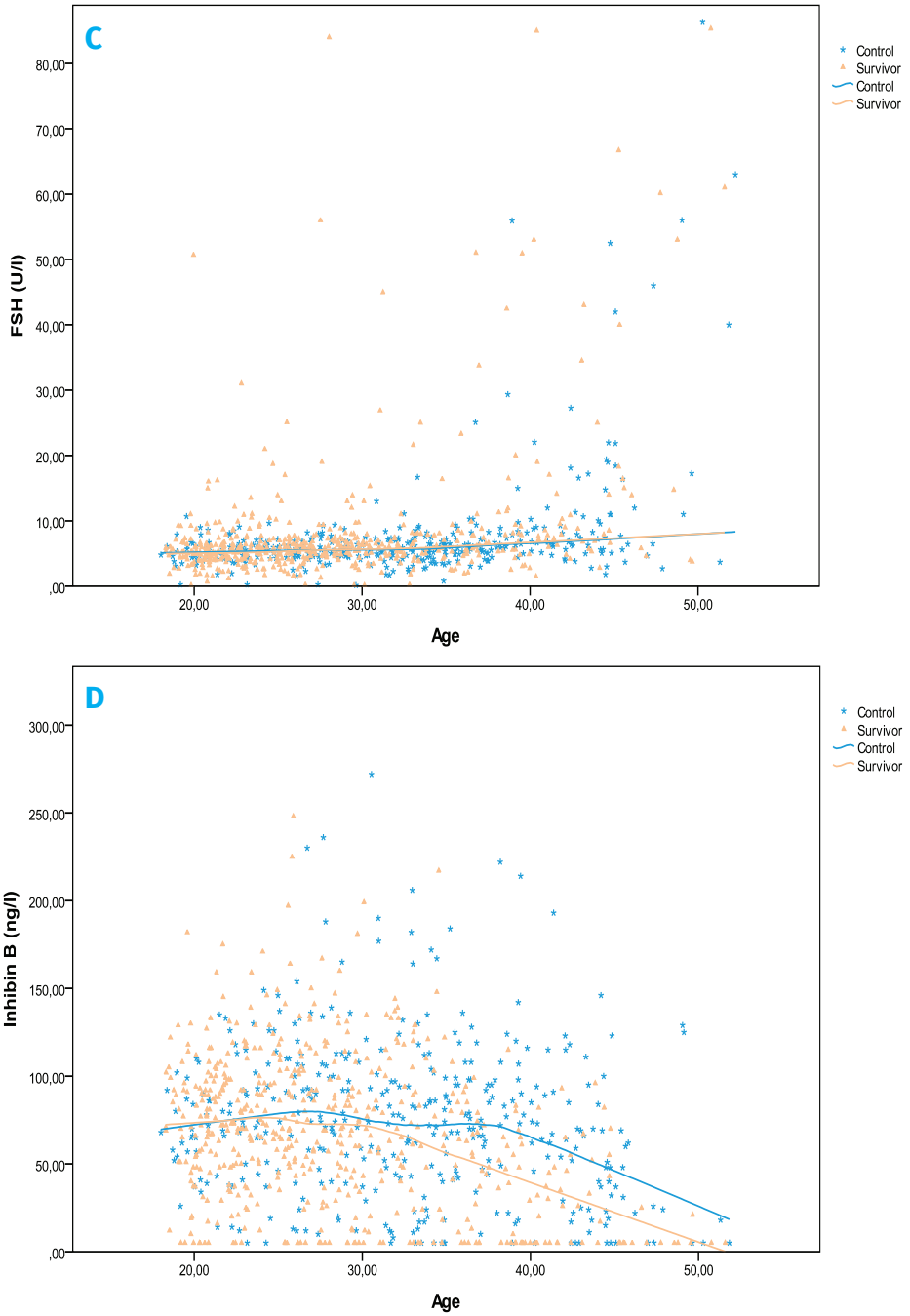


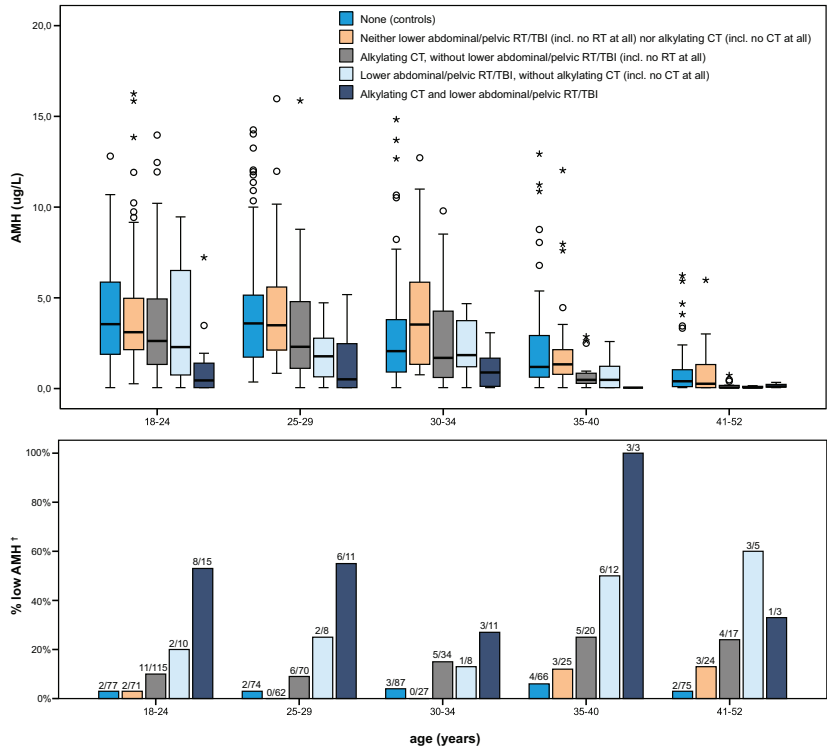
Figure 1 Levels of ovarian function markers of childhood cancer survivors and controls according to age at time of study: AMH (A), number of antral follicles (B), FSH (C), inhibin B (D)
Blue and red lines in figures represent fitted values from a locally weigthed polynomial regression (Loess) computed and plotted using SPSS.

Table 2 Study participants with low AMH levels, low AFC levels, high FSH levels, and low inhibin B levels within total group and within age subgroups

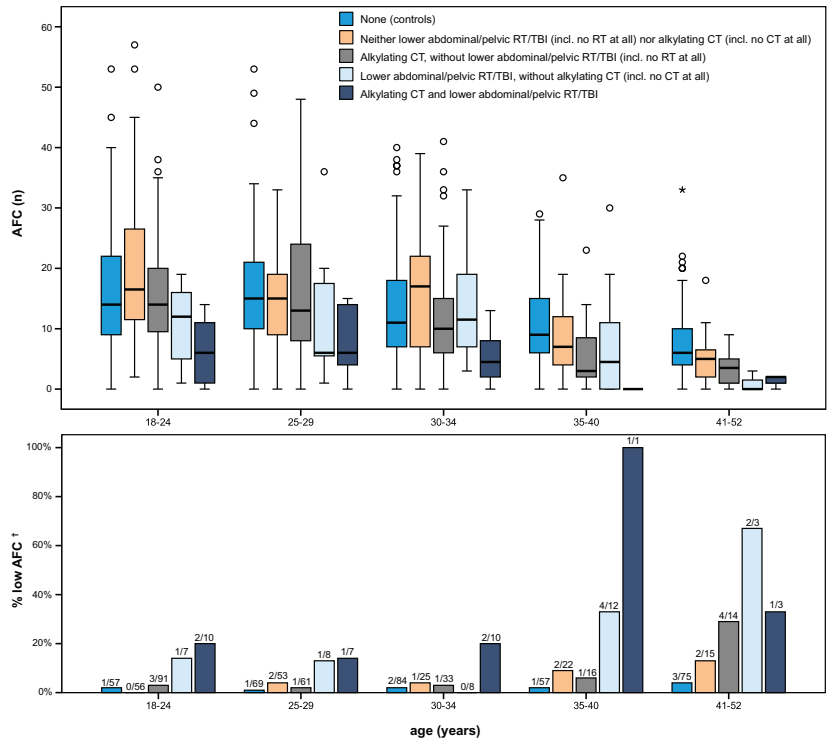
	Low AMH			Low AFC			High FSH			Low inhibin B		
	CCSs	Controls	P value	CCSs	Controls	P value	CCSs	Controls	P value	CCSs	Controls	P value
Total group	74/551 (13.4)	13/380 (3.4)	< 0.001	32/455 (7.0)	8/339 (2.4)	0.003	67/552 (12.1)	39/387 (10.1)	0.33	97/548 (17.7)	52/382 (13.6)	0.09
Age at study ≥ 18.0-24.9 yrs	23/211 (10.9)	2/77 (2.6)	0.03	6/164 (3.7)	1/57 (1.8)	0.48	16/212 (7.5)	1/80 (1.3)	0.04	19/209 (9.1)	6/78 (7.7)	0.71
Age at study ≥ 25.0-29.9 yrs	14/151 (9.3)	2/74 (2.7)	0.07	5/129 (3.9)	1/69 (1.4)	0.34	9/151 (6.0)	0/75 (0)	0.03	20/151 (13.2)	7/74 (9.5)	0.41
Age at study ≥ 30.0-34.9 yrs	9/80 (11.3)	3/87 (3.4)	0.05	4/76 (5.3)	2/84 (2.4)	0.34	8/80 (10.0)	3/89 (3.4)	0.08	13/80 (16.3)	17/89 (19.1)	0.63
Age at study ≥ 35.0-39.9 yrs	17/60 (28.3)	4/66 (6.1)	0.001	8/51 (15.7)	1/57 (1.8)	0.01	12/60 (20.0)	6/66 (9.1)	0.08	20/60 (33.3)	4/66 (6.1)	< 0.001
Age at study ≥ 40-52.9 yrs	11/49 (22.4)	2/75 (2.7)	< 0.001	9/35 (25.7)	3/71 (4.2)	0.001	22/49 (44.9)	29/76 (38.2)	0.45	25/48 (52.1)	18/74 (24.3)	0.002

Values represent the number (%) of women. AFC: antral follicle count; CCSs: childhood cancer survivors. Low AMH and low AFC were defined using age-specific cut-off values, i.e. 2 SD below the mean value of the control subjects within the concerning age group. High FSH and low inhibin B were defined using fixed cut-off values (i.e. ≥ 10 U/l and < 20 ng/l, respectively). CCSs: childhood cancer survivors.

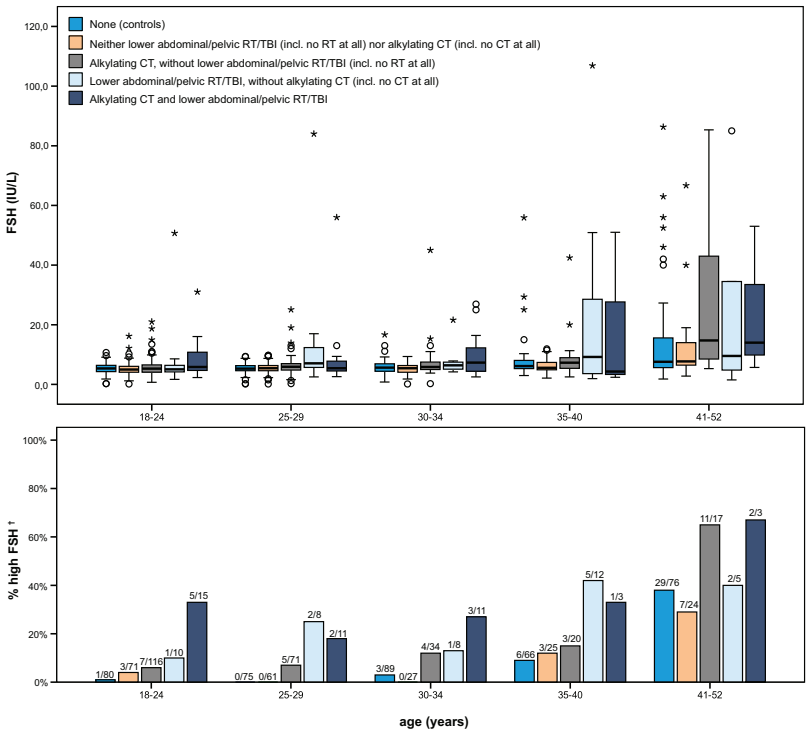
A


None (controls)
Neither lower abdominal/pelvic RT/TBI (incl. no RT at all) nor alkylating CT (incl. no CT at all)
Alkylating CT, without lower abdominal/pelvic RT/TBI (incl. no RT at all)
Lower abdominal/pelvic RT/TBI, without alkylating CT (incl. no CT at all)
Alkylating CT and lower abdominal/pelvic RT/TBI
AMH (ug/L)
20,0
15,0
10,0
5,0
0,0
18-24
25-29
30-34
35-40
41-52
% low AMH †
100%
80%
60%
40%
20%
0%
2/77 2/71 11/115 2/10 8/15
2/74 0/62 6/70 2/8 6/11
3/87 0/27 5/34 1/8 3/11
4/66 3/25 5/20 6/12 3/3
2/75 3/24 4/17 3/5 1/3
age (years)

B


None (controls)
Neither lower abdominal/pelvic RT/TBI (incl. no RT at all) nor alkylating CT (incl. no CT at all)
Alkylating CT, without lower abdominal/pelvic RT/TBI (incl. no RT at all)
Lower abdominal/pelvic RT/TBI, without alkylating CT (incl. no CT at all)
Alkylating CT and lower abdominal/pelvic RT/TBI
AFC (n)
60
50
40
30
20
10
0
18-24
25-29
30-34
35-40
41-52
% low AFC †
100%
80%
60%
40%
20%
0%
1/57 0/56 3/91 1/7 2/10
1/69 2/53 1/61 1/8 1/7
2/84 1/25 1/33 0/8 2/10
1/57 2/22 1/16 4/12 1/1
3/75 2/15 4/14 2/3 1/3
age (years)

C


None (controls)
Neither lower abdominal/pelvic RT/TBI (incl. no RT at all) nor alkylating CT (incl. no CT at all)
Alkylating CT, without lower abdominal/pelvic RT/TBI (incl. no RT at all)
Lower abdominal/pelvic RT/TBI, without alkylating CT (incl. no CT at all)
Alkylating CT and lower abdominal/pelvic RT/TBI
FSH (IU/L)
% high FSH †
1/80 3/71 7/116 1/10 5/15
0/75 0/61 5/71 2/8 2/11
3/89 0/27 4/34 1/8 3/11
6/66 3/25 3/20 5/12 1/3
29/76 7/24 11/17 2/5 2/3
18-24
25-29
30-34
35-40
41-52
age (years)

D

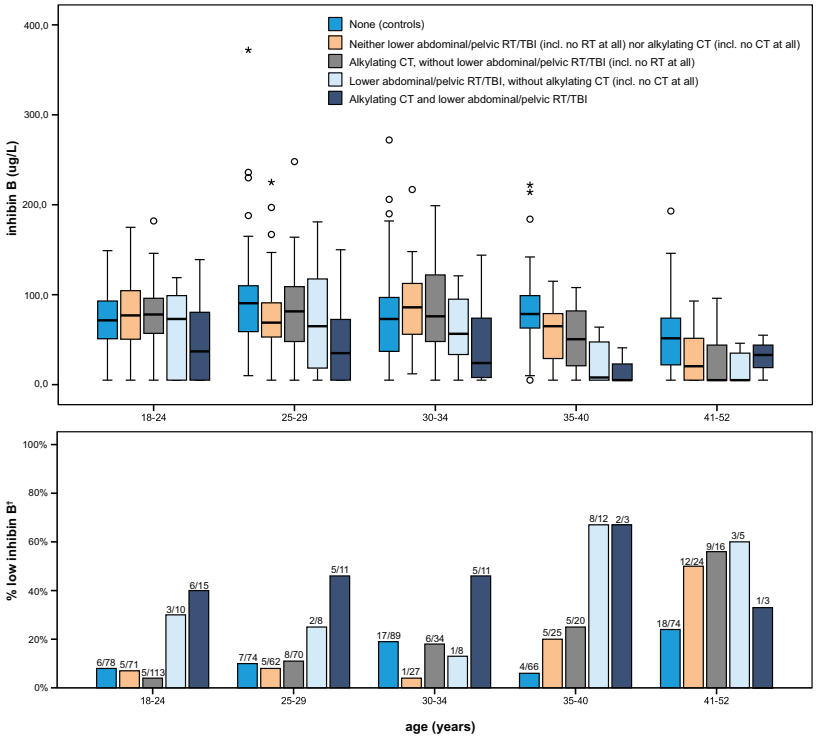

None (controls)
Neither lower abdominal/pelvic RT/TBI (incl. no RT at all) nor alkylating CT (incl. no CT at all)
Alkylating CT, without lower abdominal/pelvic RT/TBI (incl. no RT at all)
Lower abdominal/pelvic RT/TBI, without alkylating CT (incl. no CT at all)
Alkylating CT and lower abdominal/pelvic RT/TBI
inhibin B (ug/L)
% low inhibin B†
6/78 5/71 5/113 3/10 6/15
7/74 5/62 8/70 2/8 5/11
17/89 1/27 6/34 1/8 5/11
4/66 5/25 5/20 8/12 2/3
18/74 12/24 9/16 3/5 1/3
18-24
25-29
30-34
35-40
41-52
age (years)

Figure 2 Distribution of ovarian function markers of childhood cancer survivors and controls according to presumed gonadotoxicity of treatment stratified by age group: AMH (A), number of antral follicles (B), FSH (C), inhibin B (D)

* Alkylating CT = treatment with busulfan, carmustine, cyclophosphamide, chlorambucil, ifosfamide, lomustine, melphalan, procarbazine, temozolomide, chlormetine, ACNU, or thiotepa.

† Low AMH and low AFC were defined using age-specific cut-off values, i.e. 2 SD below the mean value of the control subjects within the concerning age group. High FSH and low inhibin B were defined using fixed cut-off values (i.e. ≥ 10 U/L and < 20 ng/l, respectively).

For AMH (borderline) significant differences were also found in the younger age groups and for FSH in all but the highest age groups. Multivariable logistic regression analyses for the overall effect of treatment showed that compared to controls CCSs were at increased risk of low AMH and high FSH levels (OR 6.5 (95% CI 1.3 to 32.5) and 14.3 (95% CI 2.4 to 86.4), respectively), but not of low AFC and inhibin B levels (OR 4.7 (95% CI 0.4 to 54.6) and 0.8 (95% CI 0.3 to 2.5), respectively).

Effect of specific treatments

Figure 2 shows the effect of broad treatment categories on the level of each ovarian function marker (upper part), and on the proportion of women with an abnormal ovarian function marker within these categories (lower part), stratified by age at study.

Combined treatment of alkylating CT and gonadotoxic RT, administered to 8% of CCSs, resulted in the lowest AMH, AFC, and inhibin B levels, and highest FSH levels in all age groups. Moreover, among the group of CCSs aged 18-34 years at study and treated with alkylating CT and gonadotoxic RT 46%, 19%, 27%, and 43% had low AMH, low AFC, high FSH, and low inhibin B levels, respectively, while the proportions of similarly aged CCSs with abnormal ovarian function markers were at most 17% in those treated with alkylating CT only and 8% in those who received no gonadotoxic treatment at all.

The effects of specific treatments on ovarian function, according to three different multivariable logistic regression models, are shown in Table 3. Model 1 shows that the risk of a reduced ovarian function is most significantly increased in CCSs who received lower abdominal/pelvic RT or TBI (gonadotoxic RT) in the past, with risk being highest when combined with alkylating CT. In these patients a significantly reduced ovarian function was shown by three out of four or by all markers, respectively. CCSs who received alkylating CT without gonadotoxic RT also demonstrated signs of a reduced ovarian function. However, this was less pronounced and was demonstrated by only two out of four ovarian function markers.

Model 2 shows that high AAD-scores (≥ 4) were significantly associated with a reduced ovarian function according to all markers except AFC (Table 3-model 2a), whereas high CED-scores were significantly associated with low AMH and high FSH levels only (Table 3-model 2b).

Finally, model 3 describes the results of the logistic regression analyses evaluating dose-effect relationships. First, it was investigated which CT agents or RT body sites were dichotomously (ever or never received) associated with these markers.

Table 3 Multivariable logistic regression analysis of low AMH levels, low AFC levels, high FSH levels, and low inhibin B levels according to three different models: (1) broad treatment subgroups based on their presumed gonadotoxicity; (2a) AAD-score; (2b) CED-score; (3) administered dose categories of individual chemotherapeutic agents and radiotherapy body sites

Risk factor		Low AMH (> 2SD below age-specific mean of control population)	
		OR (95% CI)	P trend[†]
Model 1 - Broad treatment subgroups*			
Type of treatment group	n		
None (controls)	390	1 (Ref)	n.a.
Neither lower abdominal/pelvic RT/TBI (incl. no RT at all) nor alkylating CT (incl. no CT at all)	214	1.9 (0.3 to 11.5)	
Alkylating CT, without lower abdominal/pelvic RT/TBI (incl. no RT at all)	262	5.8 (1.1 to 32.1)	
Lower abdominal/pelvic RT/TBI, without alkylating CT (incl. no CT at all)	44	24.1 (3.8 to 150.5)	
Alkylating CT and lower abdominal/pelvic RT/TBI	44	46.4 (7.9 to 273.3)	
Model 2a - AAD score[$]			
AAD-score	n		
Zero	650	1 (Ref)	< 0.001
1	64	1.9 (0.6 to 5.9)	
2	97	0.9 (0.3 to 2.9)	
3	86	2.1 (0.8 to 5.3)	
≥ 4	55	8.3 (3.4 to 19.8)	
Model 2b - CED score[$]			
CED-score, mg/m^2	n		
Zero	650	1 (Ref)	0.001
> 0 to < 4000	132	1.3 (0.5 to 3.5)	
≥ 4000 to 8000	78	2.1 (0.8 to 5.5)	
≥ 8000	90	4.7 (2.1 to 10.7)	
Model 3 - Dose categories of individual chemotherapeutic agents and radiotherapy body sites[†]			
Chemotherapeutic agents	n		
Cyclophosphamide, mg/m^2			
Zero	736	1 (Ref)	0.12
1-3000	65	2.3 (0.8 to 6.6)	
> 3000-4800	83	1.4 (0.4 to 4.4)	
> 4800	71	2.2 (0.8 to 5.6)	

Low AFC (> 2SD below age-specific mean of control population)		High FSH (≥ 10 U/l)		Low InhB (< 20 ng/l)	
OR (95% CI)	P trend†	OR (95% CI)	P trend†	OR (95% CI)	P trend†
1 (Ref)	n.a.	1 (Ref)	n.a.	1 (Ref)	n.a.
2.4 (0.2 to 32.7)		5.0 (0.7 to 36.23)		0.5 (0.1 to 1.6)	
3.2 (0.3 to 41.7)		12.6 (1.9 to 83.6)		0.6 (0.2 to 1.8)	
17.4 (1.2 to 243.3)		36.2 (4.8 to 274.4)		2.1 (0.6 to 7.8)	
14.6 (1.1 to 193.9)		51.4 (7.2 to 366.6)		3.6 (1.0 to 12.2)	
1 (Ref)	0.39	1 (Ref)	< 0.001	1 (Ref)	0.01
0.6 (0.1 to 5.0)		0.5 (0.1 to 4.3)		0.6 (0.2 to 1.9)	
0.7 (0.1 to 3.6)		1.1 (0.4 to 3.5)		0.8 (0.3 to 2.0)	
1.0 (0.3 to 3.7)		2.7 (1.1 to 6.7)		1.5 (0.7 to 3.1)	
2.0 (0.5 to 7.1)		6.1 (2.3 to 16.0)		2.8 (1.2 to 6.1)	
1 (Ref)	0.63	1 (Ref)	0.03	1 (Ref)	0.41
0.8 (0.2 to 3.7)		0.5 (0.1 to 2.0)		0.9 (0.4 to 2.0)	
1.3 (0.3 to 4.9)		2.2 (0.8 to 5.9)		2.0 (0.9 to 4.4)	
1.2 (0.4 to 4.2)		4.3 (1.9 to 9.8)		1.4 (0.7 to 2.9)	
--		--		--	
--		--		--	
--		--		--	
--		--		--	

Table 3 - Continued

Risk factor		Low AMH (> 2SD below age-specific mean of control population)	
		OR (95% CI)	P trend†
Model 3 - Dose categories of individual chemotherapeutic agents and radiotherapy body sites†			
Procarbazine, mg/m²			
Zero	905	1 (Ref)	< 0.001
1-2800	18	3.7 (0.8 to 15.9)	
> 2800-4600	15	12.4 (3.0 to 52.0)	
> 4600	16	15.3 (4.3 to 53.7)	
Other gonadotoxic AA§			
No	940	1 (Ref)	n.a.
Yes	14	19.7 (4.3 to 89.6)	
Mercaptopurine, mg/m²			
Zero	756	--	
1-10186	56	--	
>10186-12785	37	--	
>12785	50	--	
Dactinomycin, mg/m²			
Zero	818	1 (Ref)	0.03
1-7.40	44	1.2 (0.4 to 3.8)	
> 7.40-12.15	46	1.3 (0.4 to 4.9)	
> 12.15	44	3.6 (1.2 to 11.0)	
Doxorubicin, mg/m²			
Zero	727	--	
1-100	80	--	
> 100-180	72	--	
> 180	75	--	
Mitoxantrone, mg/m²			
Zero	941	--	
1-48	6	--	
> 48	7	--	
Dacarbazine mg/m²			
Zero	933	--	
1-1500	13	--	
> 1500	8	--	

Low AFC (>2SD below age-specific mean of control population)		High FSH (≥ 10 U/l)		Low InhB (< 20 ng/l)	
OR (95% CI)	P trend†	OR (95% CI)	P trend†	OR (95% CI)	P trend†
1 (Ref)	0.94	1 (Ref)	< 0.001	1 (Ref)	0.10
1.7 (0.3 to 10.4)		3.1 (0.5 to 20.7)		3.2 (0.9 to 11.2)	
1.6 (0.1 to 21.4)		13.4 (2.3 to 76.9)		3.3 (0.7 to 14.5)	
1.3 (0.1 to 13.2)		22.8 (5.3 to 97.6)		5.3 (1.2 to 22.9)	
1 (Ref)	n.a.	1 (Ref)	n.a.	--	
1.6 (0.1 to 24.5)		40.1 (9.1 to 177.4)		--	
1 (Ref)	0.87	--		--	
0¶		--		--	
1.3 (0.2 to 10.0)		--		--	
1.1 (0.1 to 12.5)		--		--	
--		--		--	
--		--		--	
--		--		--	
--		--		--	
1 (Ref)	0.62	--		--	
0.2 (0.02 to 2.3)		--		--	
3.0 (0.6 to 16.5)		--		--	
1.1 (0.3 to 3.9)		--		--	
--		1 (Ref)	0.99	--	
--		0¶		--	
--		1.5 (0.1 to 33.2)		--	
--		1 (Ref)	0.34	--	
--		0.2 (0.02 to 1.9)		--	
--		0.4 (0.04 to 4.6)		--	

Table 3 - Continued

Risk factor		Low AMH (> 2SD below age-specific mean of control population)	
		OR (95% CI)	P trend†
Model 3 - Dose categories of individual chemotherapeutic agents and radiotherapy body sites†			
Radiotherapy body sites			
Spinal, Gy			
Zero	920	1 (Ref)	0.21
1-23.4	12	3.1 (0.6 to 17.0)	
> 23.4-34.9	10	1.8 (0.2 to 19.1)	
> 34.9	11	2.1 (0.2 to 19.9)	
Abdomen and/or pelvis, Gy			
Zero	911	1 (Ref)	< 0.001
1-20	16	8.2 (2.0 to 33.7)	
> 20-30	13	11.0 (2.7 to 44.9)	
> 30	13	6.6 (1.3 to 32.5)	
TBI, Gy			
No	939	1 (Ref)	n.a.
Yes	14	∞ (>1 to ∞)#	
Other, Gy			
Zero	919	--	
1-28.5	11	--	
> 25.8-50.1	12	--	
> 50.1	11	--	

low AMH and low AFC were defined using age-specific cut-off values, i.e. 2 SD below the mean value of the control subjects within the concerning age group. High FSH and low inhibin B were defined using fixed cut-off values (i.e. ≥ 10 U/l and < 20 ng/l, respectively). CT: chemotherapy; RT: radiotherapy; TBI: Total Body Irradiation. All analyses were adjusted for age at study, time since diagnosis, pubertal status at treatment, smoking status, current body mass index (BMI), and hormonal contraceptive use at time of clinical measurements.

** Alkylating CT = treatment with busulfan, carmustine, cyclophosphamide, chlorambucil, ifosfamide, lomustine, melphalan, procarbazine, temozolomide, chlormetine, ACNU, or thiotepa.*

† P-value trend: p-value for test of trend for total group of study participants (based on continuous dose variable).

‡ Model 1 includes those CT agents and RT sites that appeared to be significant in the multivariable regression model using dichotomous variables (ever vs. never treated).

§ Other gonadotoxic alkylating agents included busulfan (n=2), melphalan (n=7), chlorambucil (n=2), lomustine (n=3); total n=14 pts.

¶ None of the patients treated with low-dose mercaptopurine had low AFC values; none of the patients treated with low-dose mitoxantrone had high FSH values.

 # All patients treated with TBI had low AMH values.

$ The AAD-score was calculated by adding the tertile score (1, 2, or 3) for each of the alkylating agents given to a particular CCS [321]. An AAD-score of zero was assigned to non-exposed CCSs. The CED-score was calculated using a specific equation developed by Green et al. [109]. Model 2 and 3 were additionally adjusted for RT abdomen/pelvis and TBI.

Low AFC (> 2SD below age-specific mean of control population)		High FSH (≥ 10 U/l)		Low InhB (< 20 ng/l)	
OR (95% CI)	P trend†	OR (95% CI)	P trend†	OR (95% CI)	P trend†
--		--		--	
--		--		--	
--		--		--	
--		--		--	
1 (Ref)	< 0.001	1 (Ref)	0.001	1 (Ref)	< 0.001
3.3 (0.5 to 22.2)		25.1 (7.0 to 89.6)		4.7 (1.4 to 15.4)	
9.3 (1.8 to 47.4)		10.1 (2.5 to 40.9)		7.2 (2.1 to 24.5)	
14.9 (2.2 to 102.4)		5.3 (0.9 to 31.2)		8.4 (2.2 to 31.6)	
1 (Ref)	n.a.	1 (Ref)	n.a.	1 (Ref)	n.a.
92.3 (12.4 to 688.7)		44.9 (9.9 to 203.7)		58.6 (11.9 to 287.5)	
--		--		1 (Ref)	0.21
--		--		1.7 (0.3 to 9.1)	
--		--		3.8 (0.9 to 16.5)	
--		--		0.8 (0.1 to 4.4)	

Cyclophosphamide, procarbazine, a composite group of "other alkylating agents" (i.e. busulfan, melphalan, chlorambucil and lomustine), dactinomycin, doxorubicin, mitoxantrone, spinal RT, abdominal/pelvic RT, TBI and the group of other RT were multivariably significantly associated with one or more abnormal ovarian function markers (Table S6.1). These CT agents and RT body sites were subsequently included in the dose-effect analyses. The risk of low AMH and high FSH levels appeared to increase with increasing doses of procarbazine, while for AFC and inhibin B such clear dose-effect relationships were not found. Furthermore, higher doses of abdominal/pelvic RT resulted in increasing risks of low AFC and inhibin B levels, while this type of RT resulted in low AMH and high FSH levels, irrespective of dose. Finally, high-dose dactinomycin was significantly associated with low levels of AMH, while low and medium doses were not. The results of the corresponding linear regression analyses are shown in Table S6.2.

Associations between ovarian function markers

Pearson correlation coefficients between AMH-AFC, AMH-FSH, and AMH-inhibin B were r=0.76, r=-0.52, and r=0.63, respectively, for CCSs and r=0.65, r=-0.46, and r=0.50 for controls. Among CCSs with low AMH levels, the proportions of CCSs with low AFC, high FSH, and low inhibin B levels were 34%, 51%, and 62%, respectively, and for controls 18% , 39% and 54% . When categorized by age, proportions of concordant combinations (i.e. low AMH with abnormal other marker) within the group of CCSs were lowest in the youngest age group for all three combinations (Table S7.1.1). This trend was not observed in controls (Table S7.1.2). Conversely, among CCSs with normal AMH levels, the proportions of CCSs with normal AFC, FSH, and inhibin B were 97%, 94%, and 89%, respectively, with similar proportions in controls. The proportions of concordant combinations (i.e. normal AMH with normal other marker) were lowest in the oldest age group for all three combinations, both in CCSs and controls (Table S7.2.1 and S7.2.2).

DISCUSSION

This is the first study to evaluate the impact of childhood cancer treatment on various clinical markers of ovarian function in a large cohort of CCSs and controls. Overall, the study is reassuring in that it shows that ovarian function after childhood cancer treatment appears unaffected in the majority of CCSs, even in those receiving presumed gonadotoxic chemotherapy only. Next to confirming that procarbazine, busulfan, melphalan, chlorambucil, lomustine, lower abdominal radiotherapy, and TBI are highly gonadotoxic, independent of the marker used to evaluate ovarian function, the study yields several novel findings.

First, the proportion of young CCSs (< age 35 years) with abnormal ovarian function was remarkably low (7.0-17.7%, depending on the marker used), even in case they had been treated with alkylating CT (2.7-8.8%). While this can be reassuring for the majority of female CCSs, the finding that the proportion of women with reduced ovarian function increases steadily and more rapidly compared to controls after the age of 35 years, indicates that these women should be counselled to pursue pregnancy timely as their reproductive lifespan may be shorter than anticipated. Moreover, specific groups of CCSs seem to be at high risk of a decreased ovarian function, regardless of age. These CCSs should be counselled adequately and new patients receiving such treatments should be referred to a reproductive specialist for fertility preservation counselling.

Second, the design and the size of the study population allowed, for the first time, to investigate reliable dose-effect relationships for several types of treatments. Abdominal/pelvic RT appeared to affect all ovarian function markers at almost any dose, but a clear dose-effect relationship was found when ovarian function was assessed by AFC and inhibin B. For procarbazine increasing doses were associated with increasing risks of low AMH and high FSH values. These evident dose-effect results are important for the design of future childhood oncology protocols in which the curative effect of the treatment is balanced with the risk of gonadotoxicity. Moreover, they can feed into clinical guidelines regarding surveillance of late effects for CCSs [324 325].

Third, no clear effect of cyclophosphamide on ovarian function was found. A reduced ovarian function after treatment with cyclophosphamide was shown only by AMH, although not in a dose-dependent way. Cyclophosphamide administered during childhood may, therefore, not be as detrimental for ovarian function as previously thought [45]. Our finding is consistent with a recent study by Thomas-Teinturier et al. and may be due to overall low cumulative doses of cyclophosphamide administered to CCSs, which indeed was the case in 85% of our population (< 10 g/m^2) [57]. In addition, the dosing schedule may have differed: chemotherapy given in a low dose over a prolonged period may have a different effect on the ovaries than high doses given over a short period of time.

Fourth, we found several discrepancies between abnormal levels of the ovarian function markers. In the general population, AMH and AFC seem to be highly correlated [294]. In our population correlation coefficients between AMH and AFC were also highest. Lack of a gold standard for ovarian function assessment prevented us from calculating performance measures such as sensitivity and specificity. Instead, reciprocal relationships between AMH and the other three markers were investigated. Results show that in many CCSs low AMH levels were not associated with abnormal AFC, FSH or inhibin B values, particularly in the younger age groups. This suggests that AMH starts to decline early in the sequence of events leading to menopause, long before changes in other markers occur.

This study has some unique assets compared to the previous studies on ovarian function in CCSs. First, we invited a well-defined nationwide cohort with long-term follow-up data. This benefits the generalizability and external validity of the study. Moreover, the number of control subjects evaluated by clinical assessments is unprecedented. This allowed us to calculate age-specific cut-off values of AMH and AFC and to compare these with values of CCSs. This seems more appropriate than using the same fixed cut-off value for all ages as it may facilitate a more timely detection of a reduced ovarian function. In addition, AFC was evaluated by a single observer, from stored three-dimensional ultrasound images using specialized software, thereby eliminating inter-observer error. Moreover, for AMH assessment an ultra-sensitive immunoassay was used allowing us to accurately measure AMH-levels even in the lowest ranges.

There are some limitations. First, although this is the first study clinically evaluating ovarian function in CCSs and controls on such a large scale, the number of participants still appeared to be limited for certain types of treatment. Therefore, dose-effect relationships could not be investigated for all types of treatment. Second, the proportion of clinical study participants who gave birth in the past was lower compared to the questionnaire-only participants. This could indicate that our study has been subject to some degree of participation bias, i.e. subjects who had proven to be fertile were less inclined to participate in the clinical part. However, since this trend appeared to be present in both the CCSs and comparison group, our results concerning differences between both groups can be validly interpreted. Third, treatment data did not include the total time period during which CT was administered. Therefore, we were not able to perform a time-dependent analysis, accounting for time at treatment for relapse or second malignancies. Fourth, we did not apply a correction for multiple testing since we did not focus on individual

significant results, but rather on trends, and, inevitably, multiple comparison corrections increases the probability of producing false negative results and losing statistical power. Finally, actual fertility could not be evaluated since some participants have not yet pursued pregnancy. Future research should focus on the value of ovarian function markers to predict pregnancy or age at menopause in CCSs later in life, taking into account a possible role for genetic factors [377]. This will allow clinicians to more accurately assess the remaining fertile life span of CCSs following cancer treatment, and to adequately counsel CCSs as well as future patients regarding family planning and fertility preservation.

In conclusion, although a reduced ovarian function is more frequently seen in CCSs compared to controls, it appears unaffected by cancer treatment in the majority of CCSs. However, compared to controls the proportion of CCSs with a reduced ovarian function, defined as low AMH, low AFC, high FSH, or low inhibin B levels, increases significantly above the age of 35 years. This may imply that although ovarian function appears intact long after childhood cancer treatment, CCSs remain at risk of a reduced reproductive life span. Moreover, CCSs treated with procarbazine, busulfan, melphalan, chlorambucil or lomustine, lower abdominal/pelvic radiation, or with TBI are at highest risk of a reduced ovarian function. Physicians must be aware of these effects and inform CCSs and future patients treated with these types of treatment about fertility preservation options and refer them to a reproductive specialist timely. Moreover, although AMH seems the first marker to decline with increasing age, measuring a full panel of ovarian function markers is still encouraged given the differences between these markers regarding their ability to detect a treatment effect. In addition, future prospective studies are needed to establish the value of the four ovarian function markers in predicting the chance of a future pregnancy, subfertility or a premature menopause and to evaluate the impact of newly developed treatment regimes.

SUPPLEMENTAL FILES

Supplemental File 1 – Cohort characteristics and study procedures

Eligible cohort members were selected from a cohort of patients treated for childhood cancer between 1963 and 2002 at one of the seven Dutch pediatric oncology - and stem cell transplant centres, collectively known as the Dutch Childhood Oncology Group – Long Term Effects after Childhood Cancer (DCOG-LATER). This group has developed a nationwide electronic database including patient and treatment details of all childhood cancer survivors (CCSs) in the Netherlands (DCOG-LATER database). The DCOG LATER cohort consists of those female CCSs who were treated for a malignancy or central nervous system tumour before the age of 18, who survived for at least five years after diagnosis, and who were at least 18 years at study entry (n=2,237 at database freezing date 1 January 2011) (Figure 1). We excluded 132 CCSs that fulfilled the exclusion criteria: living abroad (n=40), not being able to speak or read Dutch (n=2), mentally retarded (n=46), and being treated for second malignant neoplasm at the time of study inclusion (n=44). Those CCSs of whom contact details were unknown despite extensive tracing were considered lost to follow-up (n=119). Of CCSs that were deceased before the start of the study (n=237) only treatment data and cause of death were abstracted. One CCS had died in the peri-partum period due to pulmonary embolism.

A total of 1,749 female CCSs were eligible and thus invited for participation in the DCOG LATER-VEVO study. These women received a study package by postal mail containing extensive study information, an informed consent, a refusal form, and a questionnaire. They were asked to complete the questionnaire and return it with a signed informed consent form. Furthermore, they were asked to indicate on the informed consent form in which parts of the study they were willing to participate. When after three weeks no response was received, subjects were sent a postal reminder. If again no response was received, the subjects were contacted by telephone asking them if they were willing to participate. Participants were not financially compensated for participation in the study. However, in case they had to visit the hospital, the associated travel expenses were reimbursed.

From the total eligible cohort 1,106 (63%) consented to participate (classified as participants). Those that did not respond to the invitation (n=272), those that indicated that they did not want to participate (n=343), and those that had previously indicated that they were not willing to participate in research (n=28) were classified as non-participants.

To determine the potential for introducing bias into the studied cohort, we compared demographic and cancer-related characteristics among participants, nonparticipants, those lost to follow-up and those deceased (Supplemental Table 1). We found that the participants were similar to the non-participants with regard to age at study (or for those lost to follow-up, the age at which the cohort was assembled), age at diagnosis, and diagnosis. The type of treatment given differed between participants and non-participants. The non-participants did not differ from those lost-to-follow up (data not shown). The deceased were more likely to have been diagnosed with CNS tumours and bone tumours, and were more likely to have been treated with chemotherapy and radiotherapy or radiotherapy alone.

Control groups

To allow for comparisons with a population that had not been treated for cancer, two types of control groups were used: sisters of participating CCSs and controls from the general population.
Sisters of CCSs were recruited by asking all participating CCSs if they had any sisters aged 18 years or older. Subsequently, those who had at least one eligible sister were asked for permission to contact and invite their sister(s) for participation in the control group of the study. A total of 352 out of 580 (61%) of the participating CCSs who had a sister gave permission to invite them for the study. This resulted in 506 sisters registered, of whom 474 were eligible and 371 participated.
General practitioners of CCSs participating in the DCOG LATER-VEVO study received a letter in which they were asked to help recruit women for the control group. Where the general practitioner was willing to co-operate, we asked them to select all women from their patient population who had the same year of birth (plus or minus 2 years, with a minimum age of 18 years) as the CCSs in their patient population. All eligible women were then sent a letter signed by the general practitioner in which the study was briefly explained. This letter was accompanied by a response form and a pre-stamped envelope. On this form, women could indicate whether they were interested in the study and wanted to receive the extensive study information package. If consent was received, the study procedures were comparable to those of CCSs and sibling controls. In total, 2,120 women were recruited from the general population by their GP's of whom 727 consented to receiving the study information. Of those, 448 women participated.
Basic characteristics as well as several fertility related characteristics were compared in both control group, described in detail earlier [391]. The participation rate of sisters was much higher than control participants from the general population (74% versus 21%, respectively), whereas considerably more effort was involved in recruiting controls from the general population. Participants in this group were significantly older and more highly educated than sister controls ($P < 0.001$ for both groups). No significant differences were observed between both types of control groups in several fertility-related characteristics, suggesting minimal bias owing to selective participation.
Data collected for CCSs and controls are the same, with the exception of data related to the anti-cancer treatment in the past.

Participant groups

CCSs as well as controls could consent to the questionnaire only or to the questionnaire combined with a clinical visit. Since data presented in the current manuscript mainly regards those women consenting to the clinical part, it is important to assess whether differences exist between the questionnaire-only-group and the clinical-group. With regard to possible selection bias, we evaluated the differences in characteristics between those participating in the questionnaire part only and those participating in both the questionnaire and the clinical part (blood and/or ultrasound) of the study (Supplemental Table 2).

Figure S1 Flow chart of childhood cancer survivors (CCSs), sister controls, and general practitioner (GP) controls invited for the study.

Table S1.1 Characteristics of participating, non-participating, lost to follow-up, and deceased survivors of childhood cancer (CCSs)

	Participating CCSs (n=1,106)
Age at study, yrs; median (IQR)	27.8 (22.6 – 34.2)
≥ 18.0-24.9 yrs	400 (36.2)
≥ 25.0-29.9 yrs	275 (24.9)
≥ 30.0-34.9 yrs	178 (16.1)
≥ 35.0-39.9 yrs	131 (11.8)
≥ 40 yrs	122 (11.0)
Diagnosis	
Leukemias, myeloproliferative diseases and myelodysplastic diseases	386 (35.4)
Lymphomas and reticuloendotherlial neoplasms	176 (16.2)
CNS and miscellaneous intracranial and intraspinal neoplasms	113 (10.4)
Neuroblastoma and other peripheral nervous cell tumours	68 (6.2)
Retinoblastoma	5 (0.5)
Renal tumours	124 (11.4)
Hepatic tumours	7 (0.6)
Bone tumours	70 (6.4)
Soft tissue tumours and other extraosseous sarcomas	75 (6.9)
Germ cell tumours, trophoblastic tumours and neoplasms of the gonads	45 (4.1)
Other malignant epithelial neoplasms and malignant melanomas	19 (1.7)
Other and unspecified malignant neoplasms	1 (0.1)
Age at diagnosis, yrs; median (IQR)	6.4 (3.3 – 11.6)
0-4 yrs	437 (40.1)
5-9 yrs	295 (27.1)
10-14 yrs	279 (25.6)
15-18 yrs	78 (7.2)
Treatment	
CT-only	554 (50.2)
RT-only	80 (7.2)
CT + RT	380 (34.4)
Other	90 (8.2)

** Values represent the number (%) of women, unless indicated otherwise. IQR = Interquartile range. For some variables total N might not correspond with the number mentioned in the heading of the table because of missing data.*

Non-participating CCSs (n=643)	CCSs lost to follow-up (n=119)	Deceased CCSs (n=237)	P value Part. vs. nonpart.	P value Part. vs. lost to follow-up	P value Part. vs. deceased
27.2 (21.3-34.0)	26.0 (21.2-32.3)	19.8 (13.2-27.6)	0.20	0.03	< 0.001
259 (40.7)	52 (44.4)	158 (66.7)			
126 (19.8)	26 (22.2)	31 (13.1)			
113 (17.7)	18 (15.4)	18 (7.6)			
81 (12.7)	14 (12.0)	11 (4.6)			
58 (9.1)	7 (6.0)	19 (8.0)			
			0.46	0.57	< 0.001
195 (32.6)	33 (29.5)	71 (30.0)			
78 (13.0)	20 (17.9)	16 (6.8)			
75 (12.5)	17 (15.2)	56 (23.6)			
40 (6.7)	5 (4.5)	4 (1.7)			
6 (1.0)	1 (0.9)	3 (1.3)			
71 (11.9)	12 (10.7)	10 (4.2)			
3 (0.5)	2 (1.8)	0			
51 (8.5)	8 (7.1)	31 (13.1)			
39 (6.5)	4 (3.6)	25 (10.5)			
26 (4.3)	7 (6.3)	8 (3.4)			
14 (2.3)	3 (2.7)	12 (5.1)			
0	0	1 (0.4)			
6.5 (3.1 – 11.1)	5.7 (2.9 – 10.8)	7.8 (3.8 – 12.8)	0.95	0.63	0.09
245 (41.0)	49 (43.8)	81 (34.2)			
169 (28.3)	31 (27.7)	64 (27.0)			
143 (23.9)	23 (20.5)	65 (27.4)			
41 (6.9)	9 (8.0)	27 (11.4)			
			0.004	0.01	< 0.001
287 (45.1)	55 (47.0)	39 (18.3)			
59 (9.3)	10 (8.5)	50 (23.5)			
209 (32.9)	32 (27.4)	113 (53.1)			
81 (12.7)	20 (17.1)	11 (5.2)			

Table S1.2 Characteristics of CCSs and controls providing questionnaire data only and CCSs and controls providing both clinical and questionnaire data for the study*

	Questionnaire only CCSs (CCS_Qx) N = 542	Clinical + questionnaire CCSs (CCS_Cl) N= 564	Questionnaire only controls (C_Qx) N=390	Clinical + questionnaire controls (C_Cl) N= 429	P value CCS_Qx vs. CCS_Cl	P value C_Qx vs. C_Cl
Age at study, yrs; median (IQR)	28.6 (22.8 - 35.3)	26.9 (22.5 - 33.0)	30.3 (23.8 - 36.2)	32.5 (26.1 - 38.2)	0.003	0.003
≥ 18.0-24.9 yrs	183 (33.8)	217 (38.5)	131 (30.5)	82 (21.1)		
≥ 25.0-29.9 yrs	122 (22.5)	153 (27.1)	77 (17.9)	75 (19.3)		
≥ 30.0-34.9 yrs	96 (17.7)	82 (14.5)	81 (18.9)	89 (22.9)		
≥ 35.0-39.9 yrs	69 (12.7)	62 (11.0)	99 (23.1)	66 (17.0)		
≥ 40 yrs	72 (13.3)	50 (8.9)	41 (9.6)	77 (19.8)		
BMI, kg/m^2, median (IQR)	23.3 (21.0 - 26.7)	22.6 (20.7 - 26.3)	23.1 (21.0 - 26.1)	22.7 (21.1 - 25.7)	0.08	0.23
< 18.5	27 (5.1)	32 (5.7)	8 (1.9)	9 (2.3)		
18.50-< 25.0	317 (59.7)	346 (62.0)	283 (66.6)	259 (67.3)		
25-< 30.0	121 (22.8)	127 (22.8)	93 (21.9)	79 (20.5)		
≥ 30.0	66 (12.4)	53 (9.5)	41 (9.6)	38 (9.9)		
Level of education					0.44	0.26
Low	53 (9.9)	44 (7.8)	17 (4.0)	9 (2.3)		
Medium	333 (62.1)	349 (62.2)	182 (42.8)	181 (46.9)		
High	150 (28.0)	168 (29.9)	226 (53.2)	196 (50.8)		

Smoking status	70 (13.3)	98 (17.6)	66 (15.5)	72 (18.6)	0.05	0.26
Never had sexual intercourse					0.43	0.06
Age at study ≥ 18.0-24.9 yrs	37/183 (20.2)	44/217 (20.3)	8/131 (6.1)	19/82 (23.2)		
Age at study ≥ 25.0-29.9 yrs	11/122 (9.0)	14/153 (9.2)	5/77 (6.5)	1/75 (1.3)		
Age at study ≥ 30.0-34.9 yrs	7/96 (7.3)	7/82 (8.5)	0/81 (0)	5/89 (5.6)		
Age at study ≥ 35.0-39.9 yrs	7/69 (10.2)	6/62 (9.7)	4/99 (4.0)	2/66 (3.0)		
Age at study ≥ 40 years	7/72 (9.7)	1/50 (2.0)	0/41 (0)	3/77 (3.9)		
Given birth					.< 0.001	0.001
Age at study ≥ 18.0-24.9 yrs	29/183 (15.9)	4/217 (1.8)	13/131 (9.9)	2/82 (2.4)		
Age at study ≥ 25.0-29.9 yrs	56/122 (45.9)	20/153 (13.1)	41/77 (53.3)	5/75 (6.7)		
Age at study ≥ 30.0-34.9 yrs	68/96 (70.8)	36/82 (43.9)	55/81 (67.9)	34/89 (38.2)		
Age at study ≥ 35.0-39.9 yrs	45/69 (65.2)	37/62 (59.7)	77/99 (77.8)	31/66 (47.0)		
Age at study ≥ 40 years	41/72 (56.9)	34/50 (68.0)	35/41 (85.4)	69/77 (89.6)		
Menopause					< 0.001	0.01
Age at study ≥ 18.0-24.9 yrs	11/183 (6.0)	1/217 (0.5)	1/131 (0.7)	0/82 (0)		
Age at study ≥ 25.0-29.9 yrs	7/122 (5.7)	7/153 (4.6)	0/77 (0)	0/75 (0)		
Age at study ≥ 30.0-34.9 yrs	5/96 (5.2)	6/82 (7.3)	1/81 (1.2)	0/89 (0)		
Age at study ≥ 35.0-39.9 yrs	7/69 (10.2)	7/62 (11.3)	2/99 (2.0)	0/66 (0)		
Age at study ≥ 40 years	20/72 (27.8)	11/50 (22.0)	14/41 (34.2)	4/77 (5.2)		

** Values represent the number (%) of women, unless indicated otherwise. IQR = Interquartile range. For some variables total N might not correspond with the number mentioned in the heading of the table because of missing data.*

When comparing CCSs consenting only to completing the questionnaire (CCS-Q) to those consenting to the questionnaire and the clinical part of the study (CCS-Cl), CCS-Cl were younger, more likely never to have been pregnant, and more likely to have experienced menopause. However, BMI, level of education, smoking status and the proportion ever having had sexual intercourse were similar.
CCS-CL were also younger and lower educated than controls participating in the questionnaire and the clinical visit (C-CL). In addition, CCS-CL were more likely never to have been pregnant, and more likely to have experienced menopause than C-CL. C-CL were comparable to controls only completing questionnaire data (C-Q) with regard to age at study, education, smoking, virginity rates, but were more likely to have been pregnant and more likely to have reached menopause.

Supplemental File 2 – Laboratory assessments

Hormonal measurements

All laboratory analyses were performed by the endocrine laboratory of the VU University Medical Center Amsterdam. Serum FSH levels were analysed by an immunometric assay (Delfia, Perkin Elmer, Wallac, Turku, Finland), with a lower limit of quantitation (LLOQ) of 0.5 IU/L. The intra-assay and inter-assay coefficient of variation (CV) were 5% and 7%, respectively, at a concentration of 2 IU/L, whereas these coefficients were 3% and 6%, respectively, at a concentration > 4 IU/L. An ultra-sensitive immunoassay (pico-AMH, AnshLabs, USA) was used to measure serum AMH since we expected a significant number of the CCSs to have low AMH levels. The intra-assay CVs were 4.1%, 1.6%, 1.8%, and 2.8% as determined at concentrations of 29 pg/mL, 126 pg/mL, 304 pg/mL, and 656 pg/mL, respectively. Inter-assay CV was 2.5 % at a level of 55 pg/mL. Prior to our study the analytical performance of the pico-AMH assay was investigated by performing an extensive validation. The results of this validation, which also includes the results of the comparison between the current assay and the more commonly used Gen II ELISA AMH assay (Beckman Coulter), are reported below. Inhibin B was measured in serum by the Gen II Inhibin B ELISA (Beckman Coulter). The LLOQ was 11 ng/l and the intra-assay and inter-assay CV were < 9% and < 10% at all concentrations, respectively.

Validation of the pico-AMH assay (Anshlabs, USA)

Serum samples were drawn from 80 subjects of whom it was expected AMH levels to vary broadly (i.e. women with normal FSH levels (< 10 U/L) and a regular menstrual cycle, women with FSH levels between 10 and 20 U/L and a menstrual cycle, women with polycystic ovary syndrome (PCOS), menopausal women, and 10 men). Serum samples were stored at -20 ˚C until analysis. All samples were diluted before being analysed with the pico-AMH assay according to the instructions of the manufacturer. Results show that the intra-assay coefficients of variation (CV) as determined in serum specimens in duplicate at levels of 29 pg/mL, 126 pg/mL, 304 pg/mL and 656 pg/mL were 4.1%, 1.6%, 1.8%, and 2.8% respectively. The inter-assay CV in a serum sample containing 55 pg/mL, was 2.5% (n=7). Matrix interference was tested by comparing serum, heparin plasma and EDTA plasma, drawn at the same time in each of eight

healthy subject subjects. Results in heparin plasma were 5% higher (P=0.047) and in EDTA plasma were 4% higher (NS) compared to the paired serum samples. Samples could be diluted in a linear way (n=3: R=0.998, R=0.999 and R=0.997). Addition of AMH standard from the kit itself to serum samples led to a mean recovery of 90% (n=4). Five serum samples underwent three freeze-thaw cycles but did not show a change compared to the result obtained upon thawing the first time (99%). We performed a comparison with the Gen II assay in 67 of the above described serum samples with a range from 2-15,000 pg/mL (Anshlabs assay). The comparison, see also Figure S2.1, resulted in the following Passing & Bablok equation: AMHGen II = 1.22 AMHAnshlabs + 74 (pg/mL) (Pearson correlation of R=0.95). We did not specifically study the lower limit of quantitation (LLOQ), but given the observation that intra-assay CV at 29 pg/mL is only 4.1%, we submit that the LOQ of the Anshlabs AMH assay is about 10 times lower than the LOQ of the Gen II assay (200 pg/mL in our hands).

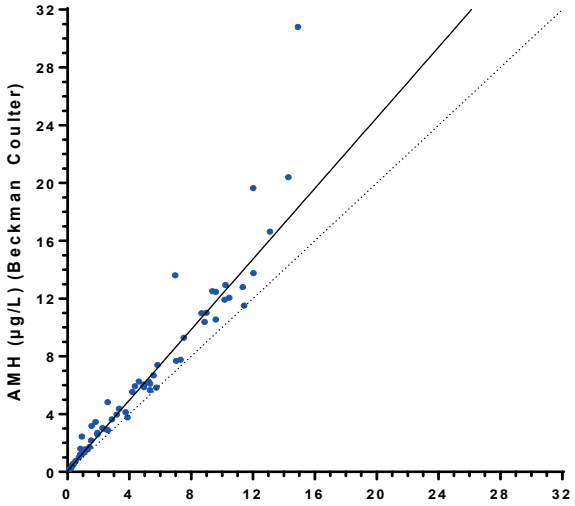


Figure S2 Comparison between the AMH Gen II assay (Beckman-Coulter) and AMH assay (Anshlabs) in 67 serum samples (Passing & Bablok analysis).

Supplemental File 3 – Ultrasound protocol

The measurements were performed using an ultrasound system with a transvaginal probe that can perform three-dimensional (3D) imaging. First, a 2D ultrasound assessment of the pelvis was performed after which an automated mechanical sweep produced the 3D data that were stored for later analysis. All 3D data were analysed by a single assessor using custom software (provided by Philips Ultrasound, Inc.) according to a pre-specified protocol. Antral follicle counts per ovary as well as the image quality of the ultrasound data were scored. The assessor analysing the 3D data had no knowledge of the previous childhood cancer diagnosis or treatment. If the ovary could not be found during the ultrasound evaluation, the antral follicle count for that ovary was imputed as zero for both 2D and 3D values.

Supplemental File 4 – Treatment details

Table S4.1 Exposure to specific chemotherapeutic agents and radiotherapy body sites for different groups of childhood cancer survivors (CCSs) (n (%))

	All participating CCSs (n=1,106)	Non-participating CCSs (n=643)	CCSs lost to follow-up (n=117)	Deceased CCSs (n=221)
Chemotherapeutic agents				
Alkylating agents				
Busulfan	6 (0.5)	8 (1.2)	0	6 2.7)
Carmustine	3 (0.3)	0	0	0
Cyclophosphamide	428 (38.7)	205 (32.0)	39 (33.3)	85 (38.1)
Chlorambucil	2 (0.2)	1 (0.2)	0	1 (0.5)
Ifosfamide	112 (10.1)	58 (9.0)	12 (10.3)	43 (19.5)
Lomustine	8 (0.7)	7 (1.1)	0	10 (4.5)
Melphalan	12 (1.1)	3 (0.5)	1 (0.9)	8 (3.6)
Procarbazine	95 (8.6)	40 (6.2)	4 (3.4)	9 (4.1)
Temozolomide	0	0	1 (0.9)	3 (1.4)
Chlormetine	74 (6.7)	31 (4.8)	5 (4.3)	4 (1.8)
ACNU	0	1 (0.2)	0	1 (0.5)
Thiotepa	1 (0.1)	1 (0.2)	0	3 (1.4)
Anti-metabolites				
Cytarabine	382 (34.6)	182 (28.4)	37 (31.6)	58 (25.8)
Methotrexate	472 (42.7)	251 (39.2)	45 (38.5)	87 (38.7)
Mercaptopurine	381 (34.5)	201 (31.4)	40 (34.2)	58 (25.8)
Thioguanine	171 (15.5)	72 (11.2)	16 (13.7)	12 (9.5)
Fluocilum	3 (0.3)	3 (0.5)	0	1 (0.5)
Hydroxycarbamide	2 (0.2)	1 (0.2)	0	1 (0.5)
Mitotic inhibitors				
Vinblastine	64 (5.8)	32 (5.0)	4 (3.4)	11 (5.0)
Vincristine	825 (74.7)	423 (66.0)	74 (63.2)	132 (58.7)
Vindesine	20 (1.8)	13 (2.0)	1 (0.9)	12 (5.4)
Etoposide	138 (12.5)	61 (9.5)	15 (12.8)	45 (20.4)
Teniposide	63 (5.7)	31 (4.8)	3 (2.6)	31 (14.0)

Table S4.1 - Continued

	All participating CCSs (n=1,106)	Non-participating CCSs (n=643)	CCSs lost to follow-up (n=117)	Deceased CCSs (n=221)
Chemotherapeutic agents				
Antitumor antibiotics				
Doxorubicin	411 (37.2)	187 (29.2)	35 (29.9)	71 (32.1)
Daunorubicin	239 (21.6)	115 (17.9)	20 (17.1)	39 (17.4)
Epirubicine	66 (6.0)	26 (4.1)	6 (5.1)	10 (4.5)
Mitoxantrone	20 (1.8)	10 (1.6)	2 (1.7)	3 (1.4)
Bleomycin	109 (9.9)	57 (8.9)	9 (7.7)	7 (3.2)
Dactinomycin	246 (22.3)	144 (22.5)	23 (19.7)	44 (19.9)
Platinum-based agents				
Carboplatin	47 (4.3)	24 (3.7)	3 (2.6)	22 (10.0)
Cisplatin	87 (7.9)	50 (7.8)	8 (6.8)	24 (10.9)
CT other				
Dacarbazine	38 (3.4)	18 (2.8)	0	3 (1.4)
Asparaginase	70 (6.3)	29 (4.5)	5 (4.3)	16 (7.2)
Peg asparaginase	3 (0.3)	2 (0.3)	1 (0.9)	2 (0.9)
Erwinase	111 (10.0)	53 (8.3)	11 (9.4)	10 (4.5)
Paronal	59 (5.3)	26 (4.1)	8 (6.8)	9 (4.1)
Crasnitin	145 (13.1)	95 (14.8)	13 (11.1)	28 (12.7)
Radiotherapy body sites				
Cranium	235 (21.3)	157 (24.6)	24 (20.3)	115 (49.8)
Spine	67 (6.1)	38 (6.0)	3 (2.5)	52 (22.5)
Thorax	82 (7.4)	41 (6.4)	7 (5.9)	34 (14.7)
Abdomen and/or pelvis	89 (8.1)	51 (8.0)	8 (6.8)	23 (10.0)
TBI	33 (3.0)	15 (2.4)	2 (1.7)	16 (6.9)
Other	69 (6.3)	35 (5.5)	8 (6.8)	29 (12.6)

Proportions do not add to 100% since patients can be included in multiple categories.

Table S4.2 Percentile distributions of cumulative doses for specific chemotherapeutic agents and radiotherapy body sites for the total group of participating childhood cancer survivors (n=1,106)

	N	Percentiles for cumulative dose *				
		10th	25th	50th	75th	90th
Chemotherapeutic agents						
Alkylating agents						
Busulfan	6	150	225	550	758	
Carmustine	3	300	300	300		
Cyclophosphamide	428	1000	2000	3150	6500	12957
Chlorambucil	2					
Ifosfamide	112	4000	6000	30000	54000	77400
Lomustine	8	352	450	600	611	
Melphalan	12	140	140	180	181	330
Procarbazine	95	1120	2100	5454	5700	8400
Temozolomide	0					
Chlormetine	74	12	32	36	72	102
ACNU	0					
Thiotepa	1	156	156	156	156	156
Anti-metabolites						
Cytarabine	382	284	649	1800	7113	26457
Methotrexate	472	1731	6000	15106	36601	46435
Mercaptopurine	381	5976	10186	12148	12786	16691
Thioguanine	171	849	849	860	975	2115
Fluocilum	0					
Hydroxycarbamide	0					
Mitotic inhibitors						
Vinblastine	64	22	36	36	72	153
Vincristine	825	8	12	21	53	71
Vindesine	20	6	6	9	12	12
Etoposide	138	600	950	1350	2160	3000
Teniposide	63	370	600	900	1305	2448
Antitumor antibiotics						
Doxorubicin	411	60	66	150	280	422
Daunorubicin	239	100	120	120	160	180
Epirubicine	66	178	240	300	400	450
Idarubicine	6	30	35	36	37	
Mitoxantrone	20	12	19	40	50	59
Bleomycin	109	36	40	75	120	180
Dactinomycin	246	3	6	10	14	23

Table S4.2 - Continued

	N	Percentiles for cumulative dose *				
		10th	25th	50th	75th	90th
Chemotherapeutic agents						
Platinum-based agents						
Carboplatin	47	920	1500	2000	3200	6051
Cisplatin	87	200	300	480	600	700
CT other						
Dacarbazine	38	885	1500	1500	3000	4500
Asparaginase, E/m²	70	43170	63140	75744	91631	120000
Pegaspariginase, E/m²	3	600	600	3000		
Erwinase, E/m²	111	40000	95000	120000	120000	160000
Paronal, E/m²	59	24000	24000	30000	80000	120000
Crasnitin, E/m²	145	64393	71122	84000	120000	153049
Radiotherapy body sites						
Cranium, Gy	235	18	24	25	50	55
Spine, Gy	67	12	18	25	35	45
Thorax, Gy	82	15	20	26	40	44
Abdomen and/or pelvis, Gy	89	10	18	25	30	40
TBI, Gy	33	6	8	8	9	12
Other, Gy	69	20	25	39	50	61

All numbers represent doses in mg/m² unless indicated otherwise.

Supplemental File 5

5a. Regression analyses procedures

Effect of specific treatments was assessed using three regression models in which different types of determinants of ovarian function were included: four broad treatment subgroups based on their presumed gonadotoxicity (model 1), accumulated scores for alkylating agent exposure (Alkylating Agent Dose (AAD) score and the Cyclophosphamide Equivalent Dose (CED) score (model 2a and 2b, respectively), and single CT agents and RT body sites (model 3). Covariates were selected as follows. First, all individual CT agents and RT body sites used in ≥ 10 CCSs were dichotomized and analysed in separate univariable linear regression models. Five previously reported gonadotoxic alkylating agents given to < 10 CCSs (busulfan, melphalan, chlorambucil, and lomustine) were accounted for by adding a dichotomous variable "Other gonadotoxic alkylating agents" to the regression model. All treatment variables with a p-value < 0.20 in univariable analyses were included in a multivariable model. The final regression model was built using a stepwise backward elimination procedure ($p < 0.05$). For all CT agents and RT sites in the final model, dose-effect analyses were performed. Doses were categorized for

each CT and RT dose variable separately using tertiles or, in case of low numbers, medians. To facilitate clinical interpretation of our data, logistic regression analyses were also performed. No variable selection was performed, but the same variables as selected in the corresponding linear regression model were used.

5b. Calculation of age-specific cut-off values used to define ovarian dysfunction

Ovarian dysfunction was defined as having AMH or AFC levels below a pre-defined cut-off value. This value was calculated using the following procedure:

1. All control subjects of the study were sorted in ascending order based on age at study.
2. All control subjects were divided into subgroups of 30 subjects, starting with first 30 subjects (i.e. the youngest group), then on to the next 30 subjects, etc.;
3. For each age subgroup the mean level of the concerning marker and the standard deviation (SD) were calculated. Since median age at menopause is 52 years in the Western World, we excluded the 19 oldest participants (n=10 controls and n=9 CCSs) .
4. Study participants were categorised as having low AMH and low AFC values if their outcome value was more than 2SD below the mean of the corresponding age control subgroup. In table S5 the age-specific cut-off values are listed per study outcome.

Table S5 Calculated age-specific cut-off values for AMH and AFC

Age (years)	n	AMH (µg/L)		AFC (n)	
		Mean	-2 SD	Mean	-2 SD
18.0 - 20.9	30	2.93	0.53	15.25	3.67
20.9 – 23.2	30	2.60	0.29	12.67	2.47
23.2 - 25.9	30	3.55	0.72	17.09	5.94
25.9 – 27.5	30	3.43	0.54	15.88	4.66
27.5- 29.1	30	2.94	0.85	12.19	3.51
29.1 - 31.6	30	2.87	0.39	15.81	3.52
31.6 - 33.2	30	1.40	0.15	9.17	2.10
33.2 – 34.9	30	1.73	0.14	12.88	2.96
34.9 – 36.5	30	2.04	0.31	13.29	5.33
36.5 – 38.7	30	1.32	0.12	9.04	1.69
38.7 – 41.8	30	0.77	0.09	9.84	4.20
41.8 – 44.5	30	0.63	0.05	6.26	1.04
44.5 – 52.2	30	0.14	0.01	5.19	1.58

Supplemental File 6 - Results of multivariable linear regression analysis

Table S6.1 Results of multivariable linear regression analysis of AMH, AFC, FSH, and inhibin B levels according to having or not ever received several chemotherapeutic agents and radiotherapy body sites

Risk factor		Change in log AMH (µg/l) β (95% CI)	Change in log AFC (no. of follicles) β (95% CI)	Change in log FSH (U/L) β (95% CI)	Change in log inhB (ng/l) β (95% CI)
Chemotherapeutic agents	n				
Cyclofosfamide					
No	736	0 (Ref)	--	--	--
Yes	219	-0.3 (-0.5 to -0.1)	--	--	--
Procarbazine					
No	905	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)
Yes	49	-1.3 (-1.6 to -0.9)	-0.4 (-0.6 to -0.1)	0.5 (0.3 to 0.8)	-0.5 (-0.8 to -0.2)
Other gonadotoxic AA*					
No	940	0 (Ref)	0 (Ref)	0 (Ref)	--
Yes	14	-1.5 (-2.1 to -0.8)	-0.5 (-1.1 to -0.04)	0.6 (0.3 to 1.0)	--
Mercaptopurine					
No	756	--	0 (Ref)	--	--
Yes	143	--	0.2 (0.1 to 0.4)	--	--
Dactinomycin					
No	818	0 (Ref)	--	--	--
Yes	134	-0.3 (-0.6 to -0.1)	--	--	--
Doxorubicin					
No	727	--	0 (Ref)	--	--
Yes	227	--	-0.2 (-0.4 to -0.1)	--	--
Mitoxantrone, mg/m²					
No	941	--	--	0 (Ref)	--
Yes	13	--	--	0.4 (0.1 to 0.8)	--
Dacarbazine mg/m²		--	--		--
No	933	--	--	0 (Ref)	--
Yes	21			-0.5 (-0.9 to -0.2)	
Radiotherapy body sites					
Spinal, Gy					
No	920	0 (Ref)	--	--	--
Yes	33	-0.5 (-0.9 to -0.04)	--	--	--
Abdomen and/or pelvis, Gy					
No	911	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)
Yes	42	-1.2 (-1.5 to -0.8)	-0.5 (-0.7 to -0.2)	0.5 (0.3 to 0.7)	-0.8 (-1.1 to -0.5)
TBI, Gy					
No	939	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)
Yes	14	-3.5 (-4.1 to -2.9)	-1.7 (-2.2 to -1.2)	0.9 (0.6 to 1.3)	0 (Ref)

** Other gonadotoxic alkylating agents included busulfan (n=2), melphalan (n=7), chlorambucil (n=2), lomustine (n=3); total n=14 patients.*

Table S6.2 Results of multivariable linear regression analysis of AMH, AFC, FSH, and inhibin B levels according to administered dose categories of chemotherapeutic agents and radiotherapy body sites

Risk factor		Change in log AMH (µg/l)	
		B (95% CI)	P trend†
Model 1 - Dose categories of individual chemotherapeutic agents and radiotherapy body sites‡			
Chemotherapeutic agents	n		
Cyclophosphamide, mg/m²			
Zero	736	0 (Ref)	0.03
1-3000	65	-0.37 (-0.67 to -0.07)	
> 3000-4800	83	-0.29 (-0.57 to -0.01)	
> 4800	71	-0.40 (-0.69 to -0.11)	
Procarbazine, mg/m²			
Zero	905	0 (Ref)	< 0.001
1-2800	18	-0.65 (-1.20 to -0.09)	
> 2800-4600	15	-1.63 (-2.21 to -1.04)	
> 4600	16	-1.55 (-2.11 to -1.00)	
Other gonadotoxic AA§			
No	940	0 (Ref)	n.a.
Yes	14	-1.52 (-2.17 to -0.87)	
Mercaptopurine, mg/m²			
Zero	756	--	
1-10186	56	--	
> 10186-12785	37	--	
> 12785	50	--	
Dactinomycin, mg/m²			
Zero	818	0 (Ref)	0.09
1-7.40	44	-0.07 (-0.41 to 0.27)	
> 7.40-12.15	46	-0.42 (-0.76 to -0.08)	
> 12.15	44	-0.59 (-0.97 to -0.20)	
Doxorubicin, mg/m²			
Zero	727	--	
1-100	80	--	
> 100-180	72	--	
> 180	75	--	
Mitoxantrone, mg/m²			
Zero	941	--	
1-48	6	--	
> 48	7	--	

Change in log AFC (no. of follicles)		Change in log FSH (U/L)		Change in log inhB (ng/l)	
B (95% CI)	P trend†	B (95% CI)	P trend†	B (95% CI)	P trend†
--		--		--	
--		--		--	
--		--		--	
--		--			
0 (Ref)	0.03	0 (Ref)	< 0.001	0 (Ref)	0.06
-0.22 (-0.61 to 0.18)		0.46 (0.12 to 0.80)		-0.48 (-0.93 to -0.03)	
-0.34 (-0.78 to 0.10)		0.58 (0.18 to 0.97)		-0.38 (-0.85 to 0.09)	
-0.52 (-0.93 to -0.11)		0.65 (0.32 to 0.98)		0.55 (-1.06 to -0.47)	
0 (Ref)	n.a.	0 (Ref)	n.a.		
-0.56 (-1.06 to -0.05)		0.63 (0.26 to 0.99)			
0 (Ref)	0.18	--		--	
0.25 (0.01 to 0.50)		--		--	
0.16 (-0.13 to 0.45)		--		--	
0.16 (-0.11 to 0.44)		--		--	
--		--		--	
--		--		--	
--		--		--	
--		--		--	
0 (Ref)	< 0.001	--		--	
-0.06 (-0.29 to 0.16)		--		--	
-0.29 (-0.53 to -0.04)		--		--	
-0.32 (-0.53 to -0.11)		--		--	
--		0 (Ref)	0.01	--	
--		0.14 (-0.41 to 0.68)		--	
--		0.75 (0.19 to 1.30)		--	

Table S6.2 - Continued

Risk factor		Change in log AMH (µg/l)	
		B (95% CI)	P trend†
Model 1 - Dose categories of individual chemotherapeutic agents and radiotherapy body sites‡			
Dacarbazine mg/m²			
Zero	933		
1-1500	13		
> 1500	8		
Radiotherapy body sites			
Spinal, Gy			
Zero	920	0 (Ref)	0.02
1-23.4	12	-0.55 (-1.21 to 0.11)	
> 23.4-34.9	10	-0.29 (-0.98 to 0.41)	
> 34.9	11	-0.78 (-1.53 to -0.04)	
Abdomen and/or pelvis, Gy			
Zero	911	0 (Ref)	< 0.001
1-20	16	-1.09 (-1.65 to -0.53)	
> 20-30	13	-1.08 (-1.69 to -0.48)	
> 30	13	-0.95 (-1.61 to -0.30)	
TBI, Gy			
No	939	0 (Ref)	n.a.
Yes	14	-3.47 (-4.09 to -2.85)	
Other, Gy			
Zero	919		
1-28.5	11		
> 25.8-50.1	12		
> 50.1	11		
Model 2 - AAD score¶			
AAD-score	n		
Zero	650	0 (Ref)	< 0.001
1	64	-0.22 (-0.53 to 0.09)	
2	97	-0.31 (-0.58 to -0.04)	
3	86	-0.58 (-0.86 to -0.30)	
≥ 4	55	-1.14 (-1.47 to -0.81)	

Change in log AFC (no. of follicles)		Change in log FSH (U/L)		Change in log inhB (ng/l)	
B (95% CI)	P trend†	B (95% CI)	P trend†	B (95% CI)	P trend†
		0 (Ref)	0.10		
		-0.78 (-1.22 to -0.34)			
		-0.20 (-0.65 to 0.25)			
0 (Ref)	0.001	0 (Ref)	0.001	0 (Ref)	< 0.001
-0.59 (-0.98 to -0.20)		0.60 (0.29 to 0.90)		-0.62 (-1.05 to -0.19)	
-0.45 (-0.87 to -0.03)		0.52 (0.18 to 0.86)		-0.84 (-1.32 to -0.37)	
-0.37 (-0.87 to 0.14)		0.20 (-0.16 to 0.57)		-0.93 (-1.44 to -0.41)	
0 (Ref)	n.a.	0 (Ref)	n.a.	0 (Ref)	n.a.
-1.73 (-2.23 to -1.22)		0.93 (0.58 to 1.28)		-2.17(-2.65 to -1.68)	
				0 (Ref)	(0.06)
				-0.45 (-1.02 to 0.11)	
				-0.58 (-1.11 to -0.05)	
				-0.10 (-0.63 to 0.43)	
0 (Ref)	< 0.001	0 (Ref)	0.03	0 (Ref)	0.01
0.04 (-0.18 to 0.27)		0.09 (-0.09 to 0.27)		0.13 (-0.12 to 0.38)	
-0.07 (-0.28 to 0.14)		0.07 (-0.08 to 0.23)		0.07 (-0.15 to 0.29)	
-0.23 (-0.43 to -0.03)		0.07 (-0.09 to 0.23)		-0.16 (-0.38 to 0.06)	
-0.57 (-0.80 to -0.33)		0.24 (0.05 to 0.43)		-0.32 (-0.58 to -0.06)	

Table S6.2 - Continued

Risk factor		Change in log AMH (µg/l)	
		B (95% CI)	P trend†
Model 3 - CED score¶			
CED-score, mg/m²	n		
Zero	650	0 (Ref)	< 0.001
> 0 to < 4000	132	-0.25 (-0.50 to -0.01)	
≥ 4000 to 8000	78	-0.66 (-0.95 to -0v36)	
≥ 8000	90	-0.84 (-1.11 to -0.57)	

** All analyses were adjusted for attained age, time since diagnosis, pubertal status at treatment, smoking status, current body mass index (BMI), and hormonal contraceptive use at time of clinical measurements. Reference groups consist of those women not exposed to the concerning type of treatment, including control subjects.*

† P-value trend for total group: p-value for test of trend for total group of study participants (based on continuous dose variable).

‡ Model 1 includes those CT agents and RT sites that appeared to significant in the multivariable regression model using dichotomous variables (ever vs. never treated).

§ Other gonadotoxic alkylating agents included busulfan (n=2), melphalan (n=7), chlorambucil (n=2), lomustine (n=3); total n=14 pts.

¶ The AAD-score was calculated by adding the tertile score (1, 2, or 3) for each of the alkylating agents given to a particular CCS [326 327]. An AAD-score of zero was assigned to non-exposed CCSs. The CED-score was calculated using a specific equation developed by Green et al.[328] Model 2 and 3 were additionally adjusted for RT adbdomen/pelvis and TBI.

Change in log AFC (no. of follicles)		Change in log FSH (U/L)		Change in log inhB (ng/l)	
B (95% CI)	P trend†	B (95% CI)	P trend†	B (95% CI)	P trend†
0 (Ref)	< 0.001	0 (Ref)	0.21	0 (Ref)	0.09
-0.03 (-0.22 to 0.15)		0.05 (-0.09 to 0.20)		0.08 (-0.12 to 0.27)	
-0.22 (-0.44 to -0.01)		0.06 (-0.11 to 0.23)		-0.16 (-0.39 to 0.08)	
-0.44 (-0.64 to -0.24)		0.21 (0.05 to 0.37)		-0.15 (-0.37 to 0.06)	

Supplemental File 7 - Associations between ovarian function markers

7.1 Childhood cancer survivors

Table S7.1.1 Numbers and proportions of concordant combinations based on low AMH levels*

		Low AFC				High FSH				Low inhibin B			
		18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs
Low AMH	18-24 yrs	5/21 24%					10/24 42%					8/24 33%				
	25-29 yrs		4/12 33%					8/14 57%					11/14 79%			
	30-34 yrs			2/7 29%					5/10 50%					8/10 80%		
	35-40 yrs				5/13 39%					9/15 60%					11/15 73%	
	40-52 yrs					5/8 63%					6/11 55%					7/10 70%

** Low AMH: > 2SD below age-specific mean of control population; Low AFC: > 2SD below age-specific mean of control population; High FSH: ≥ 10 U/L; Low inhibin B: < 20 ng/L.*

Table S7.1.2 Numbers and proportions of concordant combinations based on normal AMH values

		Normal AFC					Normal FSH					Normal inhibin B				
		18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs
Normal AMH	18-24 yrs	150/151 99%					194/201 97%					188/199 95%				
	25-29 yrs		108/110 98%				129/130 99%					120/131 92%				
	30-34 yrs			60/63 95%			65/67 97%					62/67 93%				
	35-40 yrs				31/32 97%		37/40 93%					33/40 83%				
	40-52 yrs					22/26 85%	22/38 58%					20/38 53%				

Table S7.2.1 Numbers and proportions of concordant combinations based on low AMH levels*

		Low AFC					High FSH					Low inhibin B				
		18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs
	18-24 yrs	1/2 50%					0/3 0%					2/3 67%				
	25-29 yrs		0/1 0%					0/1 0%					0/1 0%			
Low AMH	30-34 yrs			0/3 0%					0/4 0%					2/4 50%		
	35-40 yrs				1/3 33%					3/3 100%					1/3 33%	
	40-52 yrs					0/2 0%					2/2 100%					2/2 100%

** Low AMH: > 2SD below age-specific mean of control population; low AFC: >2SD below age-specific mean of control population; high FSH: ≥ 10 U/L; low inhibin B: < 20 ng/L.*

Table S7.2.2 Numbers and proportions of concordant combinations based on normal AMH values

		Normal AFC					Normal FSH					Normal inhibin B				
		18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs	18-24 yrs	25-29 yrs	30-34 yrs	35-40 yrs	40-52 yrs
	18-24 yrs	50/50 100%					76/77 99%					72/77 94%				
	25-29 yrs		66/67 99%					74/74 100%					67/73 92%			
Normal AMH	30-34 yrs			79/81 98%					79/82 96%					68/82 83%		
	35-40 yrs				53/53 100%					60/64 94%					61/64 95%	
	40-52 yrs					65/68 96%					44/69 64%					51/67 76%

11

Practice, Attitude and Knowledge of Dutch pediatric oncologists regarding female fertility (the PAK-study)

A. Overbeek
M.H. Van den Berg
L.A. Louwé
E. Wendel
M.M. ter Kuile
G.J.L. Kaspers
A.M. Stiggelbout
E. van Dulmen-Den Broeder
C. Hilders

Neth J Med.
2014 Jun;72(5):264-70.

ABSTRACT

Background
Chemotherapy and radiotherapy for childhood cancer can result in a decreased reproductive function. It is therefore important that pediatric oncologists discuss the possible impact of treatment on female fertility and available fertility preservation options with their patients. However, it is unknown what Dutch pediatric oncologists know about of the effect of cancer treatment on female fertility, whether or not they address this issue in clinical practice, what their attitudes are towards addressing fertility after cancer treatment and fertility preservation options, and to what extent they require additional information resources.

Methods
In this nationwide quantitative cross-sectional study a survey was sent to all registered pediatric oncologists in the Netherlands (n=64).

Results
Thirty-seven pediatric oncologists participated (participation rate 58%). Fertility issues were discussed with patients and/or parents by 97%. Of the pediatric oncologists, 54-76% were aware of possibilities for fertility preservation; however only < 25% reported a moderate or high confidence in their knowledge of these techniques. Pediatric oncologists stated that they had little resources to counsel their patients and 92% found educational resources not completely sufficient.

Conclusion
Pediatric oncologists are well aware of the effect that cancer treatment may have on female fertility and their responsibility to counsel their patients and/or the parents on this issue. They do not (yet) possess the knowledge to sufficiently counsel these patients and, if needed, do not frequently refer them to a fertility specialist.

INTRODUCTION

In Western countries, childhood cancer mortality rates declined by more than 50% between 1975 and 2006 as a result of more effective treatments identified and implemented during this period [329]. However, the anti-cancer treatments given to achieve these lower mortality rates may adversely affect reproductive function. In women, the pool of primordial follicles in the ovaries is fixed, and chemotherapy and radiotherapy can substantially deplete this oocyte pool. This may lead to ovarian dysfunction, infertility and premature menopause. Late effects of cancer treatment on fertility outcomes in childhood cancer survivors have been evaluated in a number of studies. Studies based on questionnaire data showed that female childhood cancer survivors had a higher risk of premature menopause [270 330-332] and were more likely to experience adverse pregnancy outcomes than their siblings due to the chemotherapy and radiotherapy these survivors received [177 178 333 334]. Recently, several studies have been conducted that measure ovarian reserve by means of anti-Müllerian hormone (AMH) or ultrasound measurements [162 335-337], showing that the ovarian reserve is indeed depleted after certain forms of chemotherapy and pelvic radiotherapy.

A nationwide cohort study on reproductive function of female childhood survivors is currently being conducted in the Netherlands (the DCOG LATER-VEVO study). Results of this study will provide insight into the effects of cancer treatment on the reproductive system of female childhood cancer survivors in the Netherlands and their risk of premature menopause. The effects of treatment in general will be assessed, as well as the effects of different treatment modalities, doses of drugs, radiation sites and doses, and age at time of treatment. The data gathered in this project will provide important information to girls with cancer (and their parents) about the possible adverse effects of treatment on the reproductive function. However, while conducting the nationwide study, it seemed that, in Dutch pediatric oncologists, knowledge about fertility issues and fertility preservation was often limited. Studies in adult oncological care indicate that knowledge about fertility issues and fertility preservation among physicians is often lacking [338-344]. In a recent study performed in Saudi Arabia, oncologists are found to have a positive attitude towards fertility preservation; however, knowledge regarding the possibilities and the success rates is poor, with up to 46% of the respondents not being familiar with any female fertility preservation options [343]. In the USA and in Canada, approximately half of the oncologists rarely referred their patients to an infertility specialist [341 344], whereas in Saudi Arabia, more than 85% did not refer [343].

Only three studies are available that have quantitatively assessed the knowledge and attitudes towards discussing female fertility issues among pediatric oncologists, two of which were performed in the USA and one in the UK [345-347]. Possibly, the lack in knowledge is due to the limited possibilities that are available in the prevention or therapy of premature menopause for female childhood cancer patients, especially when the patient is prepubertal.

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 215

Available established fertility preservation options consist of cryopreservation of embryos, vitrification of oocytes and ovarian transposition. Experimental techniques include cryopreservation of ovarian tissue, and cryopreservation of the whole ovary including vascular anastomoses. Table 1 provides a short overview of the available techniques and their limitations in female childhood cancer patients [348-350].

Table 1 Procedures and limitations of fertility preservation techniques

Technique	Procedure	Limitations
Established techniques		
Cryopreservation of embryos	Hormonal stimulation of the ovary with exogenous FSH. Ultrasound-guided transvaginal oocyte pick-up. Fertilization of the oocyte with the sperm in vitro. Primary freezing of the embryos. Embryo transfer after cancer treatment and follow-up is complete.	Not applicable to prepubertal girls Male partner or sperm donor is obligate May delay anti-cancer treatment
Vitrification of oocytes	Hormonal stimulation of the ovary with exogenous FSH. Ultrasound-guided transvaginal oocyte pick-up. Rapid freezing (vitrification) of the oocytes. Fertilization and embryo transfer after cancer treatment and follow-up is complete.	Not applicable to prepubertal girls May delay anti-cancer treatment
Ovarian transposition	Laparoscopic procedure to remove ovaries from the radiation field	Effect of chemotherapy remains Scatter radiation
Experimental techniques		
Cryopreservation of ovarian tissue	Laparoscopic or laparotomic procedure to retrieve strips of ovarian cortex. Strips are vitrified. Reimplantantion of the strips (heterotopically or orthotopically) after cancer treatment and follow-up is complete.	Success rate unknown Risk of reseeding malignancy
Transplantation of the whole ovary	Transplantation of the whole ovary with vascular anastomoses.	Success rate unknown No pregnancies reported with this method

To assess the current practice, the attitudes, and the knowledge of Dutch pediatric oncologists involved in oncological care regarding fertility and fertility preservation options in female childhood cancer patients, the PAK study was performed.

METHODS

The PAK study was designed as a nationwide quantitative cross-sectional study. Approval for the study was obtained and a waiver of informed consent was received from the Medical Ethics Committee of the VU University Medical Center.

Study population

The study population consisted of pediatric oncologists registered with the Dutch Childhood Oncology Group (DCOG, n=64). Pediatric oncologists were retrospectively excluded in case of retirement, or if they had treated less than five girls, aged 0-18 years, in the past year. The rationale to exclude these subjects was to ensure recent and adequate amount of experience with treating female pediatric patients.

Data collection

Contact information of the pediatric oncologists was provided by the DCOG. The DCOG is a collaboration between pediatric oncologists and other involved experts working in the seven pediatric oncology and stem cell transplant centres in the Netherlands. Each pediatric oncologist was sent a study information package by post. This package contained a hardcopy of the survey, a cover letter and a pre-stamped and addressed return envelope, together with login details for filling out the online version of the survey, if preferred. In addition, the pediatric oncologists were asked to fill out a refusal form if they decided not to participate. This form included several questions regarding characteristics of the pediatric oncologist as well as a question regarding the reason for not wanting to participate in the study. After three to six weeks, pediatric oncologists who had not yet responded were sent a reminder letter by post together with another copy of the study information package. If no response was received within three months, a reminder was sent by email. This email included a hyperlink, which could directly be followed in order to fill out the survey or the refusal form online. If the pediatric oncologist also did not respond to the reminder by email, the pediatric oncologist was considered a non-responder. Participants were not reimbursed for completed surveys.

Survey development

The survey was adapted from the survey used by Duffy et al. (2012) and was translated from English to Dutch by two independent medical translators. The two forward translations were carefully compared and a reconciled version was then back translated. The original survey was based on qualitative studies with oncologists and recommendations from a national advisory panel of experts in survivorship and reproductive technologies were incorporated [340]. It was slightly modified and some questions were deleted altogether, to account for differences in patient group (young age) and the fact that parents are often involved in decision-making regarding medical issues of their children. In general, questions regarded girls aged 0-18 years with cancer. For some questions, discrimination was made between pre- and post-pubertal girls. The survey covered issues related to female

fertility and fertility preservation in cancer treatment and included the following sections: (1) physician characteristics; (2) current practice; (3) availability and need for information or training; (4) knowledge; and (5) attitude. Five-point Likert scales were used in questions with regard to the pediatric oncologist's attitude and the confidence in their knowledge of fertility and fertility preservation in girls with cancer. We decided not to directly test knowledge. It was assumed that this might create a sense of an 'exam', which might lead to non-participation. However, in this way, it was not possible to report on the objective knowledge of pediatric oncologists as was done by Goodwin et al. [346]

Statistical analysis

The data were checked for normal distribution. Descriptive statistics were performed on all variables. IBM SPSS Statistics, version 20.0.0 for Windows was used for all analyses.

RESULTS

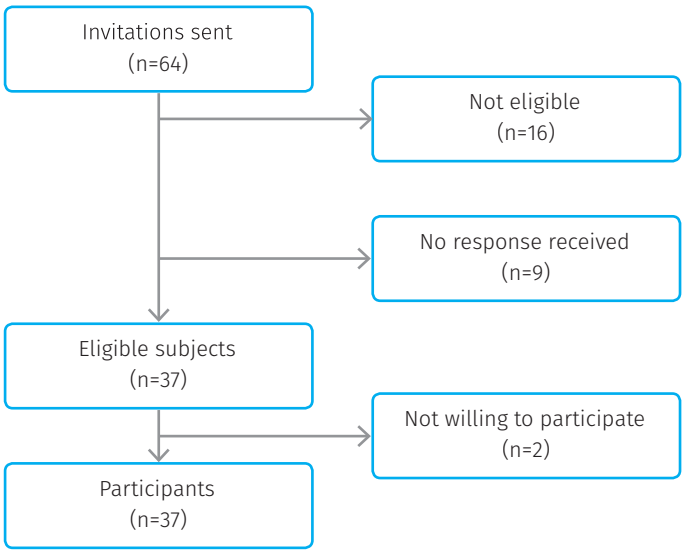


Figure 1 Flow chart depicting accrual and participation rates.

Response rate and pediatric oncologists' characteristics

In total, 64 pediatric oncologists were sent a study invitation, of whom 39 (61%) were deemed eligible. Fourteen (22%) persons were not eligible, because they had treated less than five girls in the past year, and an additional two (4%) claimed not to be eligible but did not provide a reason for this. There were nine non-responders (14%), of whom five were female and four were male. Because of the anonymous design of

the study, we were not able to evaluate whether there were differences in attitude between the participants and the non-responders. Finally, of 39 eligible subjects, 37 (58%) agreed to participate. The reasons for not participating were insufficient time (n=1) and being invited for surveys too frequently (n=1) (Figure 1).
Of the participants approximately half were male. Seventy-two per cent were between 40 and 60 years old. The median number of years in practice was 12 years (Table 2).

Table 2 Characteristics of the participating pediatric oncologists (n=37)

	Participants (n=37) **N (%)**
Sex	
Male	18 (48.6)
Female	19 (51.4)
Age	
30-39 years	9 (24.3)
40-49 years	18 (48.6)
50-59 years	9 (24.3)
> 60 years	1 (2.7)
Years of experience	
Median	12
Range	1-30

Practice

Eleven pediatric oncologists (30%) treated 5-10 children aged 0-12 years annually, whereas another 30% treated 10-20 children, aged 0-12 years. Eight oncologists treated more than 20 children aged 0-12 years annually. Seven indicated that they were not sure how many children they treat. In the age group 12-18 years, 18 oncologists treated 5-10 children, seven treated 10-20 children and three treated more than 20 children annually. In this age group, nine oncologists indicated that they did not know how many children they treated. Seventy-five per cent of the pediatric oncologists reported to usually or always discuss fertility issues before the onset of treatment with prepubertal girls or their parents and 89% discussed the issue with postpubertal girls. Almost all pediatric oncologists (97%) discussed the issue with the parents if the patient was a prepubertal girl and 32% discussed it with the girl herself. In case the girl was postpubertal, 84% of the pediatric oncologists discussed the issue with the parents and 97% with the girl herself. More than three-quarters (77%) of the pediatric oncologists indicated to spend between 5-15 minutes on fertility issues, whereas 20% spent more than 15 minutes. Approximately half of the pediatric oncologists (46%) often referred their female patients to a fertility specialist, whereas 38% sometimes referred, 3% always referred and 11% never referred.

516278-L-bw-Overbeek
Processed on: 19-1-2018

Perceived availability of fertility preservation options in own centre

All pediatric oncologists were asked which fertility preservation options were available in their own centre. As the survey was anonymous, it was not possible to substantiate the answers in the centres concerned. Therefore, when the pediatric oncologists affirmed that the requested technique was available in their centre or when they stated that it was not available, the answer was labelled 'aware of availability'. Those pediatric oncologists who responded they did not know whether that technique was available in their centre were labelled 'unaware of availability'. Most pediatric oncologists were aware of the possibilities for cryopreservation of ovarian tissue, for ovarian transposition and for embryo cryopreservation (76%, 68%, and 65%, respectively). However, it appeared that only 54% were aware of the presence of the options for oocyte cryopreservation (Table 3).

Table 3 Perceived availability of fertility preservation options in own centre (n=37)

	N (%)
Cryopreservation of ovarian tissue	
Aware of availability	28 (75.7)
Not aware of availability	8 (21.6)
Transposition of the ovaries	
Aware of availability	25 (67.6)
Not aware of availability	12 (32.4)
Cryopreservation of embryos	
Aware of availability	23 (62.2)
Not aware of availability	13 (35.1)
Cryopreservation of oocytes	
Aware of availability	19 (51.4)
Not aware of availability	17 (45.9)
Transplantation of the ovary	
Aware of availability	8 (21.6)
Not aware of availability	28 (75.7)

Percentages may not add up to 100 percent due to missing values.

Information resources for female patients

It was asked which information resources for female patients were available in each centre about fertility and fertility preservation after cancer treatment. Thirty-five per cent of pediatric oncologists stated that a printed brochure was available, and 14% reported that they had a list with references to resources with regard to fertility and fertility preservation at their disposal. Forty-one per cent reported specialised nurses or social workers trained to inform female patients about fertility issues to be available. One-third of the pediatric oncologists (30%) reported to have a fertility specialist available to refer the female patient to. Sixteen per cent of the pediatric oncologists reported there were no resources at all available for female patients.

Information and education resources for pediatric oncologists

Pediatric oncologists themselves were most likely to use the scientific literature in order to stay updated on the subject of fertility preservation (68%). Other resources used were national guidelines (35%), consult with fertility specialist (19%) or scientific meetings (5%). Three per cent of the pediatric oncologists stated that the information available on fertility preservation was not at all sufficient, while 89% found the available information to be rather or largely sufficient. Eight per cent reported the available information to be completely sufficient.

Knowledge

Overall, pediatric oncologists had a moderate or high confidence (score 4 or 5 on Likert scale) in their own knowledge of the effects of chemotherapy and radiotherapy on fertility (81% and 78% for chemotherapy and radiotherapy, respectively). However, few pediatric oncologists had a moderate or high confidence in their knowledge of ovarian transposition (24%), IVF protocols for the cryopreservation of embryos (19%) and oocytes (5%), and ovarian tissue cryopreservation (14%). Confidence in knowledge regarding health risks for the mother or foetus during pregnancy associated with various cancer treatments was rated moderate to high in 24% (mother) to 49% (foetus) of the pediatric oncologists (Table 4).

Attitude

Respondents were asked to which extent they felt it is their responsibility to discuss fertility issues with their female patients. Ninety-seven per cent reported to find it largely to entirely their responsibility to discuss infertility with the girl or parent, whereas 75% perceived it was largely or entirely their responsibility to discuss fertility preservation. In addition, pediatric oncologists were asked whether they would accept a decrease in disease-free survival in order to increase the chance of preserving fertility. Not only their own opinion on this matter was questioned, but also their judgment regarding the proportion of decrease in survival that girls and/or parents would be willing to accept. Remarkably, many pediatric oncologists (70%) did not answer these two questions. Those pediatric oncologists who did answer the question (n=11) accepted at most a 1-5% decrease in disease-free survival, and they judged that parents or patients would accept the same amount.

Perceived barriers

Pediatric oncologists were asked whether in daily clinical practice they experience certain barriers that make it less likely for them to discuss fertility or fertility preservation. The barriers reported (Table 5) were mainly related to the healthcare system, physicians' attitude, medical considerations and patient characteristics that made it less likely to discuss fertility (preservation options). Many pediatric oncologists (89%) stated that insufficient time is an important barrier to discuss fertility issues with the patients or their parents. In addition, one-third of the pediatric oncologists found their lack of knowledge about fertility preservation options a barrier. Approximately 1 in 5 pediatric oncologists reported that the lack of scientific data on the effectiveness of fertility preservation options in women

516278-L-bw-Overbeek
Processed on: 19-1-2018

Table 4 Proportion of pediatric oncologists reporting moderate or high confidence in knowledge of fertility issues and options for preservation (n=37)

Item	N (%)
The risk of infertility associated with the specific chemotherapy agents that you prescribe most often	30 (81.1)
The risk of infertility associated with abdominal and pelvic irradiation	29 (78.4)
Health risks to the foetus associated with the mother having received various cancer treatments	18 (48.6)
Health risks to the mother associated with pregnancy after various cancer treatments	9 (24.3)
Surgical techniques to protect the ovary from radiation damage	9 (24.3)
Performing current protocols for IVF cycles before cancer treatment in order to freeze embryos	7 (18.9)
Cryopreserving ovarian tissue containing primordial follicles for later auto transplantation after cancer treatment	5 (13.5)
Use of GnRH agonists prior to treatment	3 (8.1)
Cryopreserving unfertilized oocytes for future fertilization and implantation after cancer treatment	2 (5.4)
Radical trachelectomy	0

Percentages may not add up to 100 percent due to missing values.

with cancer influenced their willingness to discuss fertility and fertility preservation. A poor prognosis for long-term survival was mentioned by 24% of the pediatric oncologists as a reason not to discuss fertility issues. Other factors, for example, whether the patient has an aggressive disease and needs rapid initiation of cancer treatment, whether the patient is under the age of 16, or whether the patient appears distressed or overwhelmed about her cancer diagnosis and/or treatment, did not seem to influence the pediatric oncologist's willingness to discuss fertility and fertility preservation options (Table 5).

DISCUSSION

This is the first study assessing the practice, knowledge and attitudes towards female fertility and cancer in pediatric oncologists in the Netherlands and continental Europe. Compared with response rates from other nationwide surveys conducted among pediatric oncologists in the UK and the USA, our response rate was higher (15%)[347] or similar (68%) [345]. Our high response rate might be due to the fact that there are only a limited number of pediatric oncologists in the Netherlands and since they are all acquainted, possibly, social desirability played a role in the willingness to complete the questionnaire. When interpreting the results of our study, some limitations should be considered. Although the response rate was high, self-selection bias might have been introduced. Pediatric oncologists who were more interested in the subjects of (in)fertility and fertility preservation options

Table 5 Barriers posed to discussing fertility and fertility preservation in women with cancer (n=37).

Item	N (%)
Patient characteristics	
Has a poor prognosis for long-term survival	9 (24.3)
Appears distressed or overwhelmed about her cancer diagnosis and/or treatment	1 (2.7)
Has aggressive disease and needs rapid initiation of cancer treatment	0
Is under age 16	0
Healthcare system barriers	
Insufficient time to discuss fertility issues with patients	33 (89.2)
Lack of knowledge about fertility preservation options	12 (32.4)
Lack of availability of fertility specialists in your geographic area	4 (10.8)
Physicians' attitude barriers	
Talking about fertility after cancer gives women false hope that they will have a normal lifespan	0
Bringing up infertility is upsetting to patients	0
Bringing up infertility could make some patients decide to forego lifesaving treatments	0
Medical considerations	
Lack of data on the effectiveness of fertility preservation options in women with cancer	8 (21.6)
Chemotherapy prior to conception could increase the risk of birth defects in offspring	7 (18.9)
Discussing options for fertility preservation could delay cancer treatment	3 (8.1)
A woman treated for cancer could have health complications during a subsequent pregnancy	2 (5.4)
The hormones used in many types of fertility preservation could stimulate the growth of cancer	2 (5.4)
A pregnancy, even after successful cancer treatment, could promote cancer recurrence	0

Reported proportions represent scores 4 or 5 on the Likert scale. Percentages may not add up to 100 percent due to missing values.

were possibly more likely to discuss fertility issues with their female patients and consequently might have been more likely to participate in this study. Further, within the questions regarding barriers to discuss fertility or referral options no distinction was made between pre- and post-pubertal girls. It is likely, and has been demonstrated by Kohler et al., that pubertal status influences the pediatric oncologist's attitude and practice regarding fertility and fertility preservation [347]. Moreover, it is plausible that the pediatric oncologist's knowledge is less extensive in prepubertal girls, because few possibilities for fertility preservation exist in this patient group and these are mostly experimental. We decided not to directly test knowledge. It was assumed that this might create a sense of an 'exam', which might lead to non-participation. However, in this way, it was not possible to report on the objective knowledge of pediatric oncologists as was done by Goodwin et al. [346].

When evaluating the answers given as well as the remarks made by the participants, it seemed that some questions in the survey were considered difficult to answer or could be interpreted in various ways. For some questions, this makes it difficult to draw unambiguous conclusions.

In accordance with previous literature [345 347], our results show that pediatric oncologists frequently discuss fertility issues, but referral rates remain relatively low. The reason for this might be that the options for fertility preservation (especially in prepubertal girls) are scarce. To date the procedure of ovarian tissue cryopreservation is still experimental, but it should be realised that it might take several years to decades before these young girls will request transplantation. It is likely that the techniques that are at this moment experimental will at that time be regarded as usual care and, moreover, success rates might be much higher. Although 75% of pediatric oncologists in the PAK study were aware of the possibilities for ovarian tissue cryopreservation, only 13.5% claimed that they were confident in their knowledge regarding this technique. Other studies found similar proportions of awareness [346 347]. These results indicate that there is a lack of knowledge among pediatric oncologists regarding fertility preservation options and that there is a need for additional education. Further education for pediatric oncologists should preferably be structured in protocols or guidelines, in order to standardise fertility preservation care as much as possible in the different centres. In addition, printed brochures on the effect of cancer treatment on fertility as well as fertility preservation options (established as well as experimental) should be available for all pediatric oncologists to hand out to their patients. Good counselling and if possible, adequate action to preserve fertility will add to the future quality of life of female childhood cancer survivors.

516278-L-bw-Overbeek
Processed on: 19-1-2018

12

General discussion, clinical implications and recommendations for future research

GENERAL DISCUSSION

Because of major improvements in the treatment of childhood cancer, the numbers of long-term survivors are increasing. This increased life expectancy however, comes at a cost for many survivors. As a result of their previous anti-cancer treatment, many CCSs experience late effects. This thesis has focused on the influence of chemotherapy and radiotherapy on the female reproductive system.

In **Chapter 2** we summarized the existing literature in a systematic review on the effects of chemotherapy on the incidence of ovarian dysfunction (i.e. age at menopause, prevalence of amenorrhea, and/or elevated FSH levels). In addition, we evaluated the relationship between type and dose of chemotherapy, age at time of treatment, and time since treatment on the one hand and ovarian function on the other hand.

We found that median age at menopause was 33.5 years in young Hodgkin's lymphoma survivors and 44 years in the total group of CCS. The most important risk factors for a reduced ovarian function that were identified in more than one study were: (1) alkylating agents, specifically procarbazine and busulfan; and, (2) older age at treatment. In addition, etoposide was found to be an independent risk factor in one large well-conducted study in Hodgkin's lymphoma survivors.

Another essential finding of this review was that the methodology of the included studies often suffered from many limitations. This was caused by small and heterogeneous study populations, insufficient follow-up time, and multimodal cancer treatment (combining both chemotherapy and radiotherapy). Because synergism may occur in multimodal treatment, multivariable regression analysis alone might not sufficiently correct for other treatments given simultaneously. Few studies were sufficiently powered to assess the effects of these treatments separately. Furthermore, treatment protocols have changed extensively over the last decades, making comparisons between study groups with the same childhood cancer type difficult. In our review, most included studies have relied on predictive models derived from cross-sectional data. Because the event of premature menopause after childhood cancer treatment requires a long follow-up time, many studies used soft end-points, such as (transient) amenorrhea or ovarian reserve markers to assess ovarian function. Only studies that collect longitudinal data until menopause can reliably interpret endocrinological and ultrasound values of ovarian function, since they are stongly influenced by age. It is therefore likely that the finding in our literature review, that older age at diagnosis is associated with higher incidence of ovarian dysfunction, is due to this pitfall.

Since we concluded in our review that most studies conducted to date have methodological limitations, we aimed to minimize these flaws in the DCOG LATER-VEVO study.

First, we critically reviewed our own design as well as our control group, and reported on the pitfalls and challenges in **Chapter 3**. We formulated recommendations in order to control bias and to validate our instruments for data collection (Table 1).

Table 1 Recommendations formulated as a result of a critical appraisal of the design of our nationwide DCOG-LATER VEVO study

Recommendations
Study population
Keep non-response or loss to follow-up to a minimum
Characterize non-responders or those lost to follow-up
Control for extent and direction of bias in final data analysis
In case the number of controls is insufficient: incorporate other types of control subjects
Choose types of controls that are representative of the source population from which the study population was derived
Characterize and control for differences of potential confounders between survivors and controls
Data collection
Compare self-reported data with a more objective source, such as medical records or registries
Conduct reliability studies to account for inter- and intra- observer variation
If possible, use data collection instruments that allow for one investigator to analyze collected data (observer bias)
Assess exposed and unexposed in same manner
Blind observer to exposure status

Our first recommendation was to keep non-response or loss to follow-up to a minimum. To achieve this, we adhered to a strict protocol of reminders to make sure we contacted all patients. Due to earlier DCOG-LATER efforts, the total nation-wide cohort had already been identified and dedicated data managers had tracked many of the patients. In the final analysis of the DCOG-LATER VEVO-study we conducted a non-response analysis to control for the extent and direction of bias due to selective participation. Between participating and non-participating CCSs a few statistically significant, but not clinically relevant, differences were found. However, there appeared to be some differences between those CCS who participated in the clinical part, and those who only consented to the questionnaire. Clinical CCSs appeared significantly younger than questionnaire-only CCSs while, in contrast, clinical controls were significantly older than questionnaire-only controls. In addition, birth rate was lower among both the clinical CCSs and controls compared to questionnaire-only CCSs and controls. This could indicate that our study has been subject to some degree of participation bias, i.e. subjects who have proven to be fertile were less inclined to participate in the clinical part of the study. Overall, this could have led to an overestimation of ovarian dysfunction in our study population. However, since this trend appeared to be present in both the CCSs and control group, our results concerning differences between both groups can be validly interpreted.

Because the number of sibling controls seemed insufficient and since we preferred a control group of a similar size as the CCS group, we included an additional source of controls. These controls were recruited via general practitioners of cases and matched for age with the CCSs. The selection and comparison between the

two subgroups was discussed in **Chapter 4**. The participation rate in the sibling controls was much higher than of GP controls, whereas considerably more effort was involved in recruiting GP controls. There were differences between the control groups with regard to education and age (GP controls were significantly older and higher educated than sister controls), but no significant differences were detected regarding fertility-related characteristics, suggesting minimal bias due to selective participation.
We also evaluated the possible introduction of bias due to non-participation of the GP-controls. Participants, refusers and non-responders were compared on the following characteristics: current age, parity, age at birth of first child, having experienced fertility problems in the past, and having consulted a gynecologist for these problems. No significant differences between the response groups were present regarding current age, parity, age at birth of first child, having had fertility problems in the past, and having consulted a gynecologist in the past, although the proportion of parous women was slightly lower among participants compared to refusers.
For the validation and assessment of bias of our instruments for data collection, we conducted several methodological substudies, which are reported in **Chapters 5 to 9.**

Questionnaire data

Most studies derived from the Childhood Cancer Survivor Study [45 53], as well as other large studies (study population > 200) that chose amenorrhea as an endpoint [46 75 80 101 103], have used self-reported questionnaires to assess menopause, which may bias results. Since amenorrhea may occur for multiple reasons (pregnancy, breastfeeding, use of hormones, PCOS, surgical interventions such as hysterectomy), assuming amenorrhea is caused by a hypergonadotropic state as a consequence of previous cancer therapy might lead to a non-differential misclassification and thus a attenuation of the effect. In contrast, those women that are postmenopausal might report a regular bleeding pattern when on hormonal substitution, with a dilution of the effect as a consequence. For this reason, in the DCOG-LATER VEVO study we decided to focus on clinical ovarian function markers, rather than on questionnaire data for the evaluation of diminished ovarian function.
With regard to the self-reported pregnancy outcomes in the questionnaire, we were able to validate those by comparing records registered in the Netherlands Perinatal Registry (PRN) to self-reported outcomes. We found that self-reported pregnancy outcomes of CCSs agree well with registry data and that outcomes reported by CCSs agree better with registry data than do those of controls. Unfortunately, we were not able to validate other variables in the questionnaire, such as medical or surgical history, cycle history or medicine use.
Since we previously identified the method of invitation as a possible pitfall for introducing bias, we also compared the method of invitation (mixed invitation (paper-based together with Web-based questionnaire) vs. Web-only invitation (Web-based questionnaire only). We concluded that sending a mixed invitation rather than a web-only invitation resulted in a similar overall response as well as participation rates. However, women who were older, had a higher level of education, or were students, were more likely to have filled out the Web-based version of the questionnaire.

Ovarian function markers

To date, the perfect ovarian function test does not exist. Although elevated FSH levels are specific as to predicting imminent menopause, FSH will only be elevated shortly before the deep plunge in fecundity rates will occur. AMH and AFC share the characteristic that they gradually decrease with age, therefore allowing earlier predictions on chances of pregnancy and age at menopause. In this thesis, we critically evaluated the clinical measures of ovarian function.

AMH

We found that AMH varies significantly during the cycle in a group of regularly menstruating healthy women, and that younger women had significantly larger fluctuations in AMH levels than older women. We warned that in young patients with, commonly, a high ovarian reserve, fluctuations of AMH may have an impact on the discriminatory capability of diagnostic and predictive tests, respectively. This finding has since been replicated in other studies [351-353].
In our study we focused on the variability within one cycle, but what about long-term variability? The long-term variability in adult women in the general population has mainly been studied in cross-sectional studies, with some of them including as many as 10–15 thousand patients [258 354-357]. Overall, the studies are in good agreement that AMH declines with advancing age according to a pattern that is similar to the exponential decay of the primordial follicular pool [358 359], which is best described by a quadratic equation [258]. It was shown that of all markers, AMH can significantly aid in the prediction of menopause in the general population in a model next to age, and does so better than AFC and FSH. The prediction model failed, however, to predict extreme menopausal ages, either very young or very old. In addition, prediction intervals remained broad, rendering it unfit for use in a clinical setting [31]. There are no studies available in which AMH is measured longitudinally in a sufficiently powered group of CCSs. The decline of AMH might be different in CCSs in comparison to healthy women, due to the treatment they received. Longitudinal studies with multiple AMH measurements in which healthy women are followed throughout their reproductive life might overcome the very wide confidence intervals due to the large inter-individual differences. Subsequently, it should be validated whether the age-related rate of decline in AMH is comparable between CCSs and controls, before implementing it in screening protocols [360]. Hypothetically, the previous gonadotoxic treatment may influence the turn-over velocity of the oocyte pool or change endocrinological thresholds, inducing a different rate of atresia in CCS in comparison to population controls [51].
Of note, an unmeasurable low AMH in the general population does not exclude the possibility of pregnancy [361]. Discussion remains whether or not AMH can predict live birth. The EAGer Trial concluded that lower and higher AMH values were not associated with fecundability in unassisted conceptions in a cohort of fertile women with a history of one or two miscarriages [362]. However, Reijnders et al. found that, in a group of infertile women, ≥ 36 years of age or when showing clinical signs of diminished ovarian reserve, a very low AMH did decrease the risk of a live-birth when pregnancy did indeed occur [363]. These conflicting findings elucidate that AMH always needs to be interpreted in the context of age and (sub)fertility and that live birth rates are dependent on numerous variables other than AMH.

516278-L-bw-Overbeek
Processed on: 19-1-2018

AFC

In this thesis we described a study in which we compared antral follicle counts using real-time 2D and stored 3D data in 50 childhood cancer survivors compared to 50 healthy sibling controls, randomly drawn from the DCOG-LATER VEVO study. In terms of methodology, the use of the 3D equipment holds several benefits: it is possible to perform ultrasounds by multiple ultrasonographists and store the data, so the interpretation of the ultrasound data can be done by a single assessor, minimizing inter-observer variation. It should be noted, however, that a significant proportion of our childhood cancer survivor study population had never had sexual intercourse in comparison to healthy controls (14% vs 4%). In these women, it did not seem ethical to perform transvaginal ultrasound and therefore these women did not undergo ultrasound measurements.

Overall, we observed that with the 3D technique in comparison to the 2D technique fewer follicles were counted. It remains unclear whether this discrepancy is caused by an underestimation of the follicles in 3D-mode or an overestimation of the follicles in 2D-mode. Our data indicate that the higher the AFC values, the more likely it is that AFCs differ (either measured by two different persons, on two different time points or with two different techniques). Image quality was associated with lower antral follicle counts and higher BMI, but not with older age.

In our methological substudy we could not rule out the confounding effect of image quality on the incidence of low AFCs. Poor image quality was found significantly more often among survivors (35%) than controls (19%). This is also reflected in the relatively low between-method CCC (concordance correlation coefficient) in survivors compared to controls. We speculated that lower image quality might be caused by cancer treatment, as it is known that chemotherapy and radiotherapy can lead to ovarian atrophy, injury to the blood vessels and focal ovarian cortical fibrosis. However, image quality may also have been influenced by a higher BMI. We found in this study that the BMI of survivors was significantly higher than the BMI of controls. Van Dorp et al. found that in their study of 191 female childhood cancer survivors, AMH and follicle counts were reduced in those with obesity and insulin resistance [364]. In addition, they found that AFC and AMH were not correlated as strongly as described in the general population. However, their analyses may have been hampered by the fact that AFC values were available in only five obese women. More research should be done on image quality of ultrasonography before the test characteristics, that were defined on subfertile IVF patients, can be applied to childhood cancer survivors.

In studies evaluating the predictive value of antral follicle counts (measured in 2D) on age at menopause in the general population, AFC performs worse than AMH. AFC lost its predictive value when female age at baseline was introduced into the model [31 365]. These studies were never performed in CCSs.

FSH and inhibin B

FSH is produced by the pituitary in response to pulsatile GnRH, and is necessary for recruitment of a new cohort of follicles [323]. Historically, it has been used to detect menopausal transition. Models in which FSH is validated for menopause prediction

in the general population are scarce. In available models it seemed that FSH did no longer have any predictive value when either AMH [366] or female age [365] was added to the prediction model.
Inhibin B is secreted from granulosa cells in FSH-dependent growing follicles and levels are correlated with the number of developing antral follicles seen on ultrasonography during the early follicular phase [11]. Its levels fall in parallel with the number of ovarian antral follicles [12]. However, no specific concentration of inhibin B has been shown diagnostically discriminatory in the general population [13]. Robertson et al. proposed the FSH: inhibin B ratio as a possible alternative of measuring the menopausal transition and predicting the onset of cycle irregularity and the possibility of impending menopause in the general population [353]. However, no studies to evaluate this measure have been performed since. Very few studies have assessed the role of inhibin B for ovarian reserve in CCSs [59 130 367 368].

Which marker should be used then?

In the DCOG LATER-VEVO study the correlation coefficients between AMH and AFC were highest. Reciprocal relationships between AMH and the other three markers were investigated among the group of CCSs. Results showed that having low AMH levels in many cases was not associated with abnormal AFC, FSH or inhibin B values. However, a normal AMH was rarely associated with abnormal other markers. We showed that these discordances occurred most frequently in the youngest CCSs. The discordancies between the markers urge us to warn against the use of only one marker to assess ovarian function, until a gold standard has been firmly established. Nevertheless, our results show that AMH seems to be the first marker to be influenced by gonadotoxic therapy and thus likely to be the most sensitive marker to assess ovarian function in CCSs. Our results also show that it is more sensitive than other markers in women under 35 years of age.

Oral contraceptives

The use of oral contraceptives might influence ovarian function tests. We found overall significant decreases in FSH and inhibin B and significant increases in AMH, AFC and ovarian volume values after discontinuation of hormonal contraception in general population women. Since our research was published, several other studies on the subject have been performed. Some found no difference in markers after contraceptive use [369-371], but these studies likely suffered from methological errors and low power. In prospective studies with larger study groups, as well as in a recent RCT, especially AMH values were found to drop from 13 to 50% after a prolonged course of contraceptives [372-374]. This evidence is indicative of a suppressive effect of hormonal contraception on circulating AMH levels, at least when considering long-term use. Thus, serum AMH concentration may not retain its accuracy as predictor of ovarian function in long-term hormonal contraceptive users. For this reason, we asked all participants in the DCOG LATER-VEVO-study to refrain from contraceptives for at least two months; however, some refused. The rate of contraceptive use was 42% for childhood cancer survivors and 34% for

controls. Of CCS, 56% indicated that they were willing to temporarily refrain from contraceptives, whereas 52% of controls consented to stop. We have performed sensitivity analyses with respect to the use of contraceptives. When women using hormonal contraceptives were excluded, all multivariable results were comparable to those in which women using hormonal contraceptives were included.

Assets and novel findings of the DCOG LATER-VEVO study

In **Chapter 10** we described the main results of the DCOG LATER-VEVO study. With the DCOG LATER-VEVO study we endeavored to fill the gaps in knowledge described in the introduction of this thesis. Methodologically, the DCOG LATER-VEVO study is unique in several aspects. In contrast to other studies in the field, we invited a well-defined nationwide cohort with long-term follow-up data and assessed the ovarian function of the majority of this cohort with clinical measures, in addition to questionnaires. Furthermore, the number of population controls that were recruited for this study, and for whom also clinical measurements were available is unprecedented. This does not only benefit the results of the DCOG LATER-VEVO study, but will also yield valuable data for further research in the field of reproductive medicine. Our methods have been subject to thorough validation. All these factors minimize the amount of bias.

Overall, the results of the DCOG LATER-VEVO study showed that the proportion of CCSs (< age 35 years) with abnormal ovarian function was remarkably low (7.0-17.7%, depending on the marker used), even after treatment with alkylating CT (2.7-8.8%). Moreover, within the CCS group only 5% reported to have a self-reported premature menopause and birth rates of CCS and controls were more or less similar (except in oldest age group). However, compared to controls, the proportions of women with reduced ovarian function increase steadily and more rapidly after age 35. Therefore, we recommend that these women should be counselled to pursue pregnancy timely as their reproductive lifespan may be shorter than anticipated. Moreover, specific groups of CCSs seem to be at high risk of a decreased ovarian function, regardless of age (i.e. those treated with procarbazine, busulfan, melphalan, chlorambucil lomustine, lower abdominal/pelvic RT, or TBI). These CCSs should be counselled adequately and new patients receiving such treatments should be referred to a reproductive specialist for fertility preservation counselling.

Due to the large number of participants we were able to establish reliable dose-effect relationships for several types of treatments. Abdominal and/or pelvic RT appeared to affect all ovarian function markers at almost any dose, but a clear dose-effect relationship was found when ovarian function was assessed by AFC and inhibin B. For procarbazine increasing doses were associated with increasing risks of low AMH and high FSH values. Remarkably, we found no clear effect of cyclophosphamide on ovarian function. A reduced ovarian function after treatment with cyclophosphamide was demonstrated only by AMH, although not in a dose-dependent way. Cyclophosphamide administered during childhood may, therefore, not be as detrimental as previously thought. The (absence of) dose-response effects are important for the design of future childhood oncology protocols in which the curative effect of the treatment is balanced with the risk of gonadotoxicity.

Remaining limitations of the DCOG LATER-VEVO study

Although this study was preceded by several methological substudies and has been conducted with rigorous methods, there remain some limitations to the design with regard to the aims posed beforehand. For our DCOG LATER-VEVO study, it seemed that many chemotherapeutic agents were associated with poorer outcomes of ovarian reserve markers in a univariate model. Due to the multi-modal treatment, combining different types of chemotherapy and radiotherapy, it remains very difficult to disentangle the separate effects.

Another limitation was that the population studied and the treatments administered were so heterogeneous that division into subgroups lead to insufficient power to detect associations, resulting in very broad confidence intervals. Since this was a nationwide study, the only solution to increase the power is to develop collaborations with other countries and join efforts.

Alike many other studies that have been performed to date, the follow-up time was long, but not nearly long enough (22.9 years (median 8.2 years). More than ninety percent of our CCS had not reached menopause at time of assessment, which makes it difficult to assess the influence of time since diagnosis on this event. We have used AMH, AFC, FSH and inhibin B levels as a surrogate for ovarian function. Likewise, actual fertility could not be evaluated since a proportion of the participants has not yet pursued pregnancy.

The DCOG LATER-VEVO study was a cross-sectional study, rendering it impossible to evaluate the inter-individual variance in markers. For this, we need longitudinal assessment of ovarian function and associate these markers with age at menopause and pregnancy rates.

Of course, predicting age at menopause and actual fertility are the outcomes that a childhood cancer survivor is really interested in. Because we were not able to evaluate age at menopause reliably in our young cohort, we are not able to shed more light on the prognostic and diagnostic value of the different ovarian reserve markers with regard to premature ovarian insufficiency, one of the gaps of knowledge that was formulated in the recommendations of the IGHG. Hopefully, longitudinal follow-up of the DCOG LATER-VEVO study participants will provide this information in the future. In addition, future research should focus on the value of ovarian function markers to predict the chance of an actual pregnancy in CCSs, also taking into account a possible role for genetic factors. The DCOG LATER-VEVO study is uniquely positioned to examine this in a few years.

Lastly, other determinants of gonadal function after childhood cancer have been described, that were not assessed in the DCOG LATER-VEVO study. These include baseline gonadal function (before start of treatment), genetic factors, lifestyle factors and fertility preservation.

Baseline gonadal function

It has been shown that AMH levels are reduced in adults and young girls with newly diagnosed cancer even before the cancer treatment has started [375 376]. Since we did not have pre-diagnosis AMH levels available for the DCOG LATER-VEVO study,

this could not be tested in our study. It would be wise to initiate measurement of baseline AMH levels before treatment for childhood cancer in all childhood cancer survivors. Although its value is still uncertain, this may provide more insight into the baseline gonadal function of girls before and after treatment.

Genetic factors

One of the most powerful predictors of age at menopause is family history. Twin studies have shown that 44% to 85% of the variance in age at natural menopause is inherited [323].
Mutations in genes functioning in hormonal regulation, DNA repair, and immune function pathways, have been associated with the age at natural menopause in genome-wide association studies (GWAS) [323]. In addition, van Dorp et al. found that SNP rs 1172822 was associated with impaired ovarian reserve and lower predicted age at menopause in adult female survivors of childhood cancer, independent of what treatment they had received [377].
In addition, genes involved in drug metabolism or DNA repair may also be of influence in the amount of overall damage the chemo- and radiotherapy do in a patient's body.

Lifestyle factors

In the general population, smoking has been shown to result in lower age-specific AMH [378 379] and AFC levels [380], higher FSH levels [380] and an earlier age at menopause [381 382].
Obesity is associated with subfertility in the general population, but also in childhood cancer survivors. Obese CCSs were at greater risk of lower follicle counts and AMH levels [383]. This finding is in contrast with recent results of the St. Jude Lifetime Cohort, in which patients with a BMI ≥ 30 kg/m^2 at the time of assessment were less likely to have a diagnosis of premature ovarian failure [48]. We have corrected all results of the DCOG LATER-VEVO study for BMI as well as current smoking behavior in our regression model of treatment related effects on ovarian function to take these factors into account.

Fertility preservation

Many reports have been published over the last five years on the revolutionary development of live births after transplanted ovarian tissue. A known concern of ovarian tissue transplantation is the presence (and possible re-implantation) of malignant cells in ovarian tissue. However, other problems may also arise. It is known that ovarian reserve is slightly reduced after unilateral oophorectomy in both childhood cancer survivors [384], as well as in adult patients [385]. To our knowledge there is no information on the decline of ovarian function due to surgery with the aim of fertility preservation. It is of pivotal importance that fertility preservation is only undertaken when the risk of an impaired future fertility is certain, and that the patient will benefit, rather than be harmed by the procedure. After all, an important part of the doctor's ethical mores is 'primum non nocere'.

In **Chapter 11** we reported the results of a survey in pediatric oncologists regarding practice, attitude and knowledge with regard to fertility and fertility preservation after cancer treatment. Although pediatric oncologists seemed to be well aware of the effect that cancer treatment may have on female fertility and their responsibility to counsel their patients and/or the parents on this issue, they stated that they did not possess the knowledge to sufficiently counsel these patients and, if needed, did not frequently refer them to a fertility specialist.

CLINICAL IMPLICATIONS

From the research described in this thesis, several implications for clinical practice can be formulated. These are summarized in Table 2.

Table 2 Clinical implications from the DCOG LATER-VEVO study

All childhood cancer survivors should be counselled to pursue pregnancy timely as their reproductive lifespan may be shorter than in healthy controls, even though ovarian reserve markers may be normal before the age of 35.
CCSs treated with procarbazine, busulfan, melphalan, chlorambucil or lomustine, lower abdominal/pelvic radiation, or with TBI are at highest risk of a reduced ovarian function. Physicians must be aware of these effects and inform CCSs and future patients treated with these types of treatment about fertility preservation options and refer them to a reproductive specialist timely.
Although AMH seems the first marker to decline with increasing age, measuring a full panel of ovarian function markers is still encouraged given the differences between these markers regarding their ability to detect a treatment effect.
AMH always needs to be interpreted in the context of age and (sub)fertility and it should be noted that live birth rates are dependent on numerous variables other than AMH.

RECOMMENDATIONS FOR FUTURE RESEARCH

This was the first report of the DCOG LATER-VEVO study, showing solid results on the gonadotoxic effect of cancer treatment on ovarian function in female childhood cancer survivors. In this thesis we have focused primarily on ovarian function. However, through questionnaires, we have gathered extensive data on fertility, the need for fertility treatment and pregnancy outcomes. These data will be analyzed and presented in the near future. We furthermore aim to prospectively follow-up this cohort to collect longitudinal data. This will not only allow us to eventually reach the endpoint in which all participants have reached menopausal age, but also to report on the predictive value of endocrine and ultrasonographic markers of ovarian function on pregnancy and menopause in childhood cancer survivors and controls. We asked all women participating in the DCOG LATER-VEVO study at what age their mother had reached menopause by means of a questionnaire, since age at menopause

is highly heritable. In the planned follow-up study, (in which the incidence rate of menopause will be higher), we will be able to analyze this factor more thoroughly. Since age at menopause is highly heritable, testing of the genome should be included in future studies. In addition, future studies on childhood cancer survivors should also focus on pharmacogenetics and -genomics, since genetic variation may also influence the way and the rate at which chemotherapy and radiotherapy are metabolized and thus the amount of damage they may cause. We are currently in the process of preparing a GWAS analysis of all DCOG LATER-VEVO participants.
We recommend that longitudinal studies should be performed with significantly longer follow-up time, in analogy with the Framingham study or Women's Health Initiative [386 387]. In the meantime, pooling of the results of previously conducted studies or a meta-analysis based on individualised patient data (IPD) could be performed. IPD meta-analysis is a specific type of systematic review. Rather than extracting summary (aggregate) data from study publications or from investigators, the original research data are requested directly from the researchers responsible for each study. These data can then be re-analysed centrally and combined, if appropriate, in meta-analyses. IPD meta-analyses can improve the quality of data and the type of analyses that can be done and produce more reliable results [388]. Recently, the PanCareLIFE initiative was launched, a 5-year European programme that studies the impact of treatment regimens on the long-term health of childhood cancer survivors. PanCareLIFE will evaluate the risks of impairments in female fertility, in hearing, and in quality of life. The results of the DCOG LATER-VEVO study, and at a later time the PanCareLIFE results, in combination with the data of the large cohort studies in the USA and Europe may be the basis of an IPD meta-analysis, which will provide more robust answers on the (dose-related) effects of all types of chemotherapy on ovarian function. Nevertheless, the impact of time and ever changing therapeutic protocols should still be encountered.

IDEALLY...

So, what do clinicians really need in order to adequately counsel their patients?
First, they would need the perfect ovarian reserve test that reliably describes the entire remaining follicle pool and that is able to predict age at menopause with small confidence intervals. This will enable clinicians to better determine the remaining reproductive window, and prevent menopause-associated conditions, including osteoporosis, cardiovascular diseases, and psychosexual dysfunction, in the long run. In addition, they would be able to counsel patients at a younger age.
Second, they would need a database containing threshold doses for gonadotoxicity of chemotherapeutics agents and radiotherapy with corresponding odds ratios for premature menopause. The DCOG LATER-VEVO study has provided important new data that may be used in such an endeavour. However, longer follow-up of the DCOG LATER-VEVO study population is needed to estimate risks for premature menopause, as we have focussed on ovarian function markers for now.
Third, they would need a prediction model in which the chance of pregnancy can be calculated, taking into account age at treatment and previous treatments and

possibly genetic factors. Such a model might be a utopia, as semen quality, tubal patency and endometrial receptivity also play a large role. In the Netherlands, reproductive specialists use the Hunault model to predict the chance of spontaneous ongoing pregnancy within one year in subfertile couples [389]. This model accounts for female age, duration of subfertility, parity and sperm motility. Hypothetically, one might use an adapted Hunault model, in which a corrected age (i.e. a modelled age higher than the woman's calendar age depending on the amount of gonadotoxic chemotherapy and radiotherapy she has received) is used. Or better yet, a prognostic model in which influencing factors (such as type and dose of therapy, age at therapy, use of contraceptives, baseline gonadal function, genetic factors, and lifestyle factors) are included.

Hopefully these models can be created in the future when more research has been done and when solid evidence as to who is at severe risk of involuntary childlessness and premature menopause has been established. New patients must be offered fertility preservation options and be given tailored advice as to when to pursue their wish for children when a shorter fertile life span is evident. When a physician would know the odds of a diminished ovarian function associated with the treatment that he/she is about to prescribe, it will become easier for pediatric oncologists to counsel, and for parents and young girls to decide on fertility preservation options. From our PAK-study, it seemed that pediatric oncologists mainly felt restricted to discuss fertility preservation due to insufficient time, lack of knowledge and lack of scientific data on the effectiveness of fertility preservation options. The last two factors may be improved with the DCOG LATER-VEVO study and future studies that will follow from this large nationwide cohort study.

13

Nederlandse samenvatting

NEDERLANDSE SAMENVATTING

In Nederland werden er in 2016 586 kinderen onder de leeftijd van 18 jaar met kanker gediagnosticeerd. Gelukkig zijn er in de afgelopen 40 jaar grote verbeteringen geweest in de behandeling van kinderkanker, waardoor de overlevingskans enorm is toegenomen. We weten echter dat de behandeling van kanker tot nadelige late effecten kan leiden. Sommige van deze late effecten kunnen pas jaren nadat de behandeling is afgerond manifest worden. Een vaak voorkomend laat effect van kankerbehandeling bij meisjes is een verminderde vruchtbaarheid. In **Hoofdstuk 2** hebben we de bestaande literatuur over de effecten van chemotherapie op de incidentie van een verminderde eicelvoorraad (uitgedrukt in leeftijd ten tijde van de menopauze, prevalentie van amenorroe, en/of verhoogde FSH spiegels) samengevat. Daarbij hebben we ook gekeken naar de relatie tussen de eicelvoorraad en het type en de dosering van de chemotherapie, de leeftijd waarop de behandeling werd toegediend en de inmiddels verstreken tijd sinds de chemotherapie. De groepen die het meeste risico hadden op een slechtere eicelvoorraad waren patiënten die behandeld waren voor borstkanker, die met het middel procarbazine behandeld waren, of die hoge doseringen alkylerende middelen toegediend hadden gekregen. Een hogere leeftijd bij diagnose is een risicofactor voor een verminderde eicelvoorraad. Helaas hadden alle studies die wij hebben geïncludeerd methodologische tekortkomingen, hierdoor was het niet mogelijk om de gegevens samen te voegen.

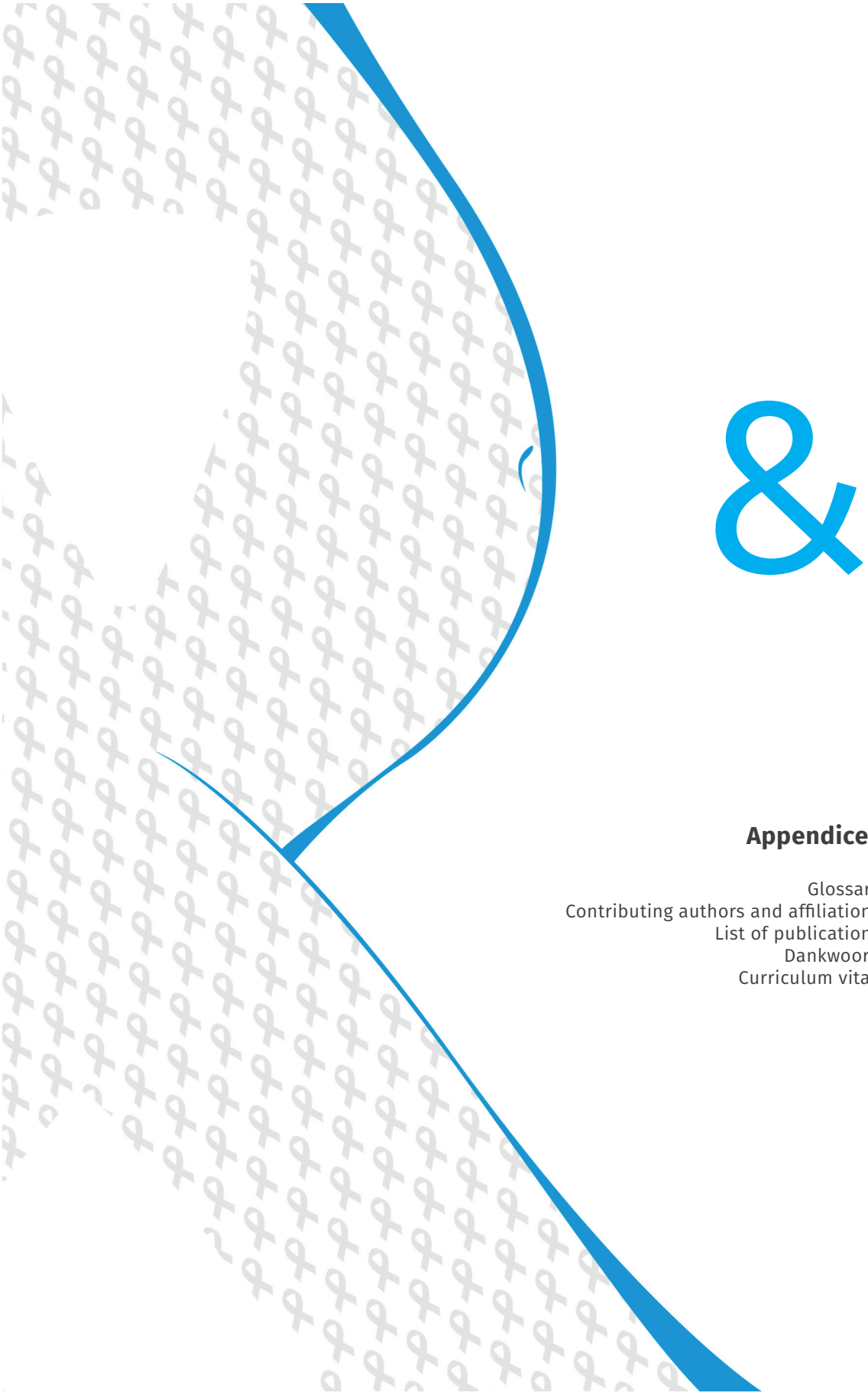
Bij het ontwerp van de DCOG LATER-VEVO studie hebben wij geprobeerd deze methodologische tekortkomingen zoveel mogelijk te voorkomen. Wij hebben ons eigen design kritisch geëvalueerd en hieruit volgden zowel valkuilen als aanbevelingen. Deze beschrijven we in **Hoofdstuk 3**. Een van de aanbevelingen was om het aantal personen dat niet reageerden op de uitnodiging voor het onderzoek tot een minimum te beperken. Dit hebben wij gerealiseerd door een strikt protocol van reminders (zowel via de post als via de telefoon) te volgen. We hebben tevens gecontroleerd of de deelnemers van het onderzoek verschilden van degenen niet wilden of konden meedoen. Als vergelijking voor de deelnemers hebben we gezonde vrouwen gevraagd dezelfde onderzoeken te ondergaan. In eerste instantie kozen wij voor zussen van de vrouwen die in hun jeugd kanker hadden gehad. Toen bleek dat de aantallen met deze strategie onvoldoende waren, hebben we een tweede controle groep uitgenodigd, bestaand uit vrouwen geworven uit de algemene populatie via huisartsen in de regio. De selectie en de vergelijking van deze twee controlegroepen wordt beschreven in **Hoofdstuk 4**. Het bleek dat het percentage dat wilde deelnemen in de zussengroep veel hoger was dan bij de vrouwen die via de huisarts warden geworven. Tevens waren er verschillen met betrekking tot het opleidingsniveau en de leeftijd (controles geworven via de huisarts waren significant ouder en hoger opgeleid dan de zussen). Er waren geen significante verschillen met het oog op vruchtbaarheid, wat suggereert dat het risico op vertekening als gevolg van selectieve deelname klein is.

In **Hoofdstuk 5** tot en met **9** beschrijven we de validatie van onze meetmethoden en evalueren we vormen van vertekening. Voor een aantal gegevens over de menstruele cyclus, hormoongebruik, zwangerschappen en leeftijd van de overgang hebben wij

voor het DCOG LATER-VEVO onderzoek vragenlijsten gebruikt. We hebben gepoogd deze vragenlijsten te valideren door de zwangerschapsgegevens te vergelijken met de data geregistreerd in de Perinatale Registratie Nederland (PRN). Het bleek dat de zelf-gerapporteerde gegevens over de zwangerschap goed overeen kwamen met de database, en dat kinderkankeroverlevenden zelfs meer accuraat rapporteerden dan de controles.
Verder hebben we gekeken of de manier van benaderen voor het onderzoek uitmaakte of vrouwen deelnamen aan het onderzoek. Een uitnodiging met zowel de optie om de vragenlijst op papier als via internet in te vullen gaf een even hoge respons als de uitnodiging voor alleen de online vragenlijst. Echter, degenen die de vragenlijst online invulden, waren vaker ouder, hoger opgeleid of nog student.
Ten aanzien van de klinische parameters (bloedmonsters en transvaginale echo) hebben we de volgende validatietoetsen gedaan. We hebben vergeleken of de waarden van FSH, LH, estradiol, anti-Müllerian hormoon (AMH), inhibine B, antrale follikel telling (AFC) en ovarieel volume, bepaald op dag 7 van de stopweek van de pil, overeen kwamen met de waarden die gemeten werden op cyclusdag 2-5 van twee opeenvolgende natuurlijke menstruele cycli. De reden hiervoor was dat veel kinderkankeroverlevenden en controles liever niet wilden stoppen met het gebruik van de anticonceptiepil voor hun deelname aan het DCOG LATER-VEVO onderzoek. Het bleek dat zowel de waarden van de hormonale bepalingen als de waarden verkregen met echo-onderzoek beïnvloed werden door het gebruik van de pil. Voor deze studie ontwikkelden we een algoritme waarmee de waarden tijdens pilgebruik konden worden omgerekend naar waarden in de natuurlijke cyclus. We hebben tevens bekeken of AMH in gezonde vrouwen fluctueert door de cyclus heen. Wat bleek is dat de fluctuaties weliswaar klein zijn en geen duidelijk patroon hebben, maar dat er toch wel aanzienlijke fluctuaties kunnen optreden, vooral bij jonge vrouwen. Deze bevinding maakt dat we voorzichtig moeten zijn om een klinische betekenis ten aanzien van de eicelvoorraad te hangen aan één bepaling van het AMH. Als derde validiteitsonderzoek hebben we bekeken hoe de vergelijkbaarheid is van de echo-resultaten geïnterpreteerd door twee beoordelaars. Hiervoor hebben we de 2D en de opgeslagen 3D data van de echobeelden in 50 kinderkankeroverlevenden en 50 gezonde zussen vergeleken. Gegevens van vrouwen die overgewicht hadden en vrouwen met een hoge AFC waren het minst goed vergelijkbaar. Er was geen duidelijk verschil tussen kinderkankeroverlevenden en controles, hoewel de overeenkomst tussen 2D en 3D in de survivor groep beduidend lager was dan in de controlegroep. Als conclusie van dit onderzoek stelden wij dat het toepassen van een geprotocolleerde 3D echoscopie de voorkeur heeft bij multicenter studies, omdat er dan een geblindeerde data analyse kan plaatsvinden door maar één beoordelaar. Echter, we konden op basis van dit onderzoek niet stellen dat 3D echo de voorkeur heeft boven de conventionele echo als het gaat om het counselen van individuele patiënten.
In **Hoofdstuk 10** hebben we de belangrijkste resultaten van de DCOG LATER-VEVO studie beschreven. Kinderkankeroverlevenden die behandeld zijn met procarbazine, met busulfan, melfalan, chlorambucil of lomustine, met abdominale of bekkenbestraling of met volledige lichaamsbestraling hadden het hoogste risico op een verminderde ovariële functie. Artsen moeten op de hoogte zijn van

deze effecten en moeten hun patiënten duidelijk en tijdig counselen over de mogelijkheden tot fertiliteitspreservatie. Zij dienen deze patiënten ook te verwijzen naar een gynaecoloog met dit aandachtsgebied. Uit onze data bleek echter ook dat bepaalde chemotherapie (waaronder cyclofosfamide), die in eerdere literatuur als zeer gonadotoxisch was aangemerkt, geen dosisafhankelijk effect had, en zelfs niet op alle markers een significant effect had.
Wat verder opviel, is dat resultaten niet per se consistent waren met betrekking tot de verschillende markers. Omdat er tot op heden geen gouden standaard is, bevelen we dus niet aan om gebaseerd op slechts één marker te counselen, maar liever om het hele panel in ogenschouw te nemen. Het leek wel alsof AMH het gevoeligst was om veranderingen op te pikken, vooral op jongere leeftijd.
Wat belangrijk is om je te realiseren, is dat de vertaling van ovariële functiemarkers, zoals AMH, AFC en FSH, naar daadwerkelijke vruchtbaarheid en vervroegde overgang nog niet gemaakt is. De meeste studies, waaronder de onze, hebben een onvoldoende lange follow-up tijd om deze markers al aan harde eindpunten te koppelen. Zelfs in de gezonde populatie is de informatie hierover schaars. Daarbij is een belangrijk punt, dat niet alleen de vroegere behandeling, maar ook genetische factoren een rol kunnen spelen bij een verminderde eicelvoorraad of een jongere leeftijd bij de menopauze. Om deze redenen bevelen wij longitudinale studies aan, waarin bij dezelfde vrouwen meerdere keren de ovariële functie wordt gemeten tot zij het eindpunt bereiken. Bij voorkeur worden in deze studie ook genetische factoren geëvalueerd.
In het laatste hoofdstuk (**Hoofdstuk 11**) bespreken wij de resultaten van een vragenlijst die wij naar kinderoncologen in Nederland hebben gestuurd. Deze vragenlijst behelsde onderwerpen over de praktijk, de attitude en de kennis van kinderoncologen met betrekking tot vruchtbaarheid na kankertherapie en fertiliteitspreservatie. Hoewel de meeste kinderoncologen goed op de hoogte waren van de schadelijke effecten van de behandeling van kanker op de vruchtbaarheid en zich ook verantwoordelijk voelden om patiënten over dit onderwerp te counselen, gaven zij aan dat ze niet voldoende kennis hadden om deze counseling adequaat uit te voeren, en dat zij meisjes en/of hun ouders onvoldoende verwezen naar een specialist op dit gebied.

516278-L-bw-Overbeek
Processed on: 19-1-2018



Appendices

Glossary
Contributing authors and affiliations
List of publications
Dankwoord
Curriculum vitae

GLOSSARY

AA: alkylating agents
AAD: alkylating agent dose
ABVD: adriamycin, bleomycin, vinblastine, dacarbazine
AFC: antral follicle count
ALL: acute lymphatic leukemia
AMH: anti-Müllerian hormone
AML: acute myeloid leukemia
ART: assisted reproductive techniques
BBT: basal body temperature
BEACOPP: bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisolone
BEAM: BCNU/carmustine, etoposide, cytarabine, melphalan
BW: birth weight
CAF: cyclophosphamide, doxorubicin, fluorouracil
CCC: concordance correlation coefficient
CCS: childhood cancer survivor
CCT: controlled clinical trial
CED: cyclophosphamide equivalent dose
CEF: cyclophosphamide, epirubicin, fluorouracil
ChlVPP: chlorambucil, vinblastine, prednisolone, procarbazine
CHOP: cyclophosphamide, doxorubicin, vincristine, prednisone
CI: confidence interval
CMF: cyclophosphamide, methotrexate, fluorouracil
CML: chronic myeloid leukemia
CT: chemotherapy
CV: coefficient of variation
DOB: date of birth
DCOG-LATER: Dutch Childhood Oncology Group – Long term effects after Childhood Cancer
DHAP: dexamethasone, cytarabine, cisplatin
E2: estradiol
EBVP: epirubicin, bleomycine, vinblastine, prednisone
FSH: follicle stimulating hormone
GA: gestational age
GnRH: gonadotropin releasing hormone
GP: general practitioner
GWAS: genome-wide association studies
HC: hormonal contraceptive
HFI: hormone free interval
HL: Hodgkin lymphoma
HRT: hormone replacement therapy
HSCT: hematopoetic stem cell transplantation
ICC: intracorrelation coefficient
ICSI: intracytoplasmatic sperm injection

IHGH: International Late Effects of Childhood Cancer Guideline Harmonization Group
IQR: interquartile range
IPD: individualized patient data
IUD: intra-uterine device
IUI: intra-uterine insemination
IVF: in vitro fertilization
LCH: Langerhans cell histiocytosis
LH: luteinizing hormone
LLOQ: lower limit of quantitation
LOA: level of agreement
LOPP: chlorambucil, vincristine, procarbazine, prednisone
MOPP: mechlorethamine, vincristine, prednisone and procarbazine
MTX: methotrexate
MVPP: mustine/nitrogen mustard, vinblastine, procarbazine and prednisolone
NC: natural cycle
NHL: non-Hodgkin lymphoma
OEPA: doxorubicin, etoposide, prednisone, vincristine
OPPA: doxorubicin, procarbazine, prednisone, vincristine
OR: odds ratio
OV: ovarian volume
PNET: primitive neuroectodermal tumor
POI: premature ovarian insufficiency
PPV: positive predictive value
PRN: Netherlands Perinatal Registry
QOL: quality of life
RCT: randomized clinical trial
RT: radiotherapy
SD: standard deviation
TAC: docetaxel, doxorubicin, cyclophosphamide
TBI: total body irradiation
VAI: vincristine, actinomycine D, ifosfamide
VIDE: vincristine, ifosfamide, doxorubicine, etoposide

CONTRIBUTING AUTHORS AND AFFILIATIONS

dr. C.M. Beerendonk, Radboud University Medical Center, Department of Obstetrics and Gynecology, Nijmegen, The Netherlands.

dr. M.H. van den Berg, VU University Medical Center, Department of Pediatrics, Subdivision Pediatric Oncology/Hematology, Amsterdam, The Netherlands

dr. D. Bresters, Willem-Alexander Children's Hospital, Leiden University Medical Center, Leiden, The Netherlands

prof. Dr. F.J. Broekmans, University Medical Center Utrecht, Department of Reproductive Medicine and Gynecology, Utrecht, The Netherlands

dr. J. van Disseldorp, University Medical Center Utrecht, Department of Reproductive Medicine and Gynecology, Utrecht, The Netherlands

dr. W. van Dorp, Sophia Children's Hospital/Erasmus MC University Medical Center, Department of Pediatric Oncology, Rotterdam, The Netherlands

dr. E. van Dulmen-den Broeder, VU University Medical Center, Department of Pediatrics, Subdivision Pediatric Oncology/Hematology, Amsterdam, The Netherlands

dr. M. Hauptmann, Netherlands Cancer Institute, Department of Epidemiology and Biostatistics, Amsterdam, The Netherlands

dr. W.J Hehenkamp, VU University Medical Center, Department of Obstetrics and Gynecology, Amsterdam, The Netherlands

dr. M. van der Heiden-van der Loo, Dutch Childhood Oncology Group, The Hague, The Netherlands

dr. A.C. Heijboer, VU University Medical Center, Department of Clinical Chemistry, Endocrine Laboratory, Amsterdam, The Netherlands

dr. M.M. van den Heuvel-Eibrink, Sophia Children's Hospital/Erasmus MC University Medical Center, Department of Pediatric Oncology, Rotterdam, The Netherlands

dr. C. Hilders, Reinier de Graaf Hospital, Department of Obstetrics and Gynecology, Delft, The Netherlands

dr. G.A. Huizinga, Beatrix Children's Hospital/University Medical Center Groningen, Department of Pediatric Oncology, Groningen, The Netherlands

drs. C.W.P.M. Hukkelhoven, The Netherlands Perinatal Registry, Utrecht, the Netherlands

prof. Dr. G.J.L. Kaspers, VU University Medical Center, Department of Pediatrics, Subdivision Pediatric Oncology/Hematology, Amsterdam, The Netherlands

dr. L.C. Kremer, Emma Children's Hospital/Academic Medical Center, Department of Pediatric Oncology, Amsterdam, The Netherlands

dr. M.M. ter Kuile, Leiden University Medical Center, Department of Obstetrics and Gynecology, Leiden, The Netherlands

prof. Dr. C.B. Lambalk, VU University Medical Center, Department of Obstetrics and Gynecology, Amsterdam, The Netherlands

Prof. Dr. J.S. Laven, Erasmus MC University Medical Center, Department of Gynecology and Obstetrics, Division Reproductive Medicine, Rotterdam, The Netherlands

prof. Dr. F.E. van Leeuwen, Netherlands Cancer Institute, Department of Epidemiology, Amsterdam, The Netherlands

dr. J.J. Loonen, Radboud University Nijmegen Medical Center, Department of Pediatric Oncology, Nijmegen, The Netherlands

drs. L.A. Louwé, Leiden University Medical Center, Department of Obstetrics and Gynecology, Leiden

dr. H.J. van der Pal, Emma Children's Hospital/Academic Medical Center, Department of Pediatric Oncology, Amsterdam, The Netherlands

dr. C.M. Ronckers, Emma Children's Hospital/Academic Medical Center, Department of Pediatric Oncology, Amsterdam, The Netherlands

dr. R. Schats, VU University Medical Center, Department of Obstetrics and Gynecology, Amsterdam, The Netherlands

prof. Dr. A.M. Stiggelbout, Leiden University Medical Center, Department of Medical Decision Making, Leiden, The Netherlands

dr. W.J.E Tissing, Beatrix Children's Hospital/University Medical Center Groningen, Department of Pediatric Oncology, Groningen, The Netherlands

prof. Dr. J.W.R. Twisk, VU University Medical Center, Department of Epidemiology and Biostatistics and the EMGO(+) Institute for Health and Care Research, Amsterdam, The Netherlands

dr. A.B. Versluys, Wilhelmina's Children's Hospital, Department of Pediatric Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

drs. E. Wendel, VU University Medical Center, Department of Pediatrics, Subdivision Pediatric Oncology/Hematology, Amsterdam, The Netherlands

drs. M.E. Wijdeveld, VU University Medical Center, Department of Obstetrics and Gynecology, Amsterdam, The Netherlands

516278-L-bw-Overbeek
Processed on: 19-1-2018

LIST OF PUBLICATIONS

Caanen MR, Schouten NE, Kuijper EAM, van Rijswijk J, van den Berg MH, van Dulmen-den Broeder E, **Overbeek A**, van Leeuwen FE, van Trotsenburg M, Lambalk CB. Effects of long-term exogenous testosterone administration on ovarian morphology, determined by transvaginal (3D) ultrasound in female-to-male transsexuals. Hum Reprod. 2017 May 13:1-8. doi: 10.1093/humrep/dex098. [Epub ahead of print]

Overbeek A, van den Berg MH, van Leeuwen FE, Kaspers GJL, Lambalk CB, van Dulmen-den Broeder E. Chemotherapy-related late adverse effects on ovarian function in female survivors of childhood and young adult cancer: A systematic review. Cancer Treat Rev. 2017 Feb;53:10-24.

Louwé LA, Stiggelbout AM, **Overbeek A,** Hilders CG, van den Berg MH, Wendel E, van Dulmen-den Broeder E, Ter Kuile MM. Factors associated with frequency of discussion of or referral for counselling about fertility issues in female cancer patients. Eur J Cancer Care (Engl). 2016 Oct 24. doi: 10.1111/ecc.12602. [Epub ahead of print]

van den Berg MH, van Dulmen-den Broeder E, **Overbeek A**, Ronckers CM, van Dorp W, Kremer LC, van den Heuvel-Eibrink MM, Huizinga GA, Loonen JJ, Versluys AB, Bresters D, Lambalk CB, Kaspers GJ, van Leeuwen FE. Fertility studies in female childhood cancer survivors: selecting appropriate comparison groups. Reprod Biomed Online. 2014 Sep;29(3):352-61.

Overbeek A, van den Berg M, Louwé L, Wendel E, ter Kuile M, Kaspers G, Stiggelbout A, van Dulmen-Den Broeder E, Hilders C. Practice, attitude and knowledge of Dutch pediatric oncologists regarding female fertility. Neth J Med. 2014 Jun;72(5):264-70.

König TE, van der Houwen LE, **Overbeek A**, Hendriks ML, Beutler-Beemsterboer SN, Kuchenbecker WK, Renckens CN, Bernardus RE, Schats R, Homburg R, Hompes PG, Lambalk CB. Recombinant LH supplementation to a standard GnRH antagonist protocol in women of 35 years or older undergoing IVF/ICSI: a randomized controlled multicentre study. Hum Reprod. 2013 Oct;28(10):2804-12.

Overbeek A, van den Berg MH, Hukkelhoven CW, Kremer LC, van den Heuvel-Eibrink MM, Tissing WJ, Loonen JJ, Versluys AB, Bresters D, Kaspers GJ, Lambalk CB, van Leeuwen FE, van Dulmen-den Broeder E; On Behalf of the DCOG LATER-VEVO Study Group. Validity of self-reported data on pregnancies for childhood cancer survivors: a comparison with data from a nationwide population-based registry. Hum Reprod. 2013 Mar;28(3):819-27.

Overbeek A, van den Berg MH, Kremer LC, van den Heuvel-Eibrink MM, Tissing WJ, Loonen JJ, Versluys B, Bresters D, Kaspers GJ, Lambalk CB, van Leeuwen FE, van Dulmen-den Broeder E. A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges. BMC Cancer. 2012 Aug 23;12(1):363.

Overbeek A, Broekmans FJ, Hehenkamp WJ, Wijdeveld ME, van Disseldorp J, van Dulmen-den Broeder E, Lambalk CB. Intra-cycle variability of anti-Müllerian hormone in individuals: a re-analysis. Reprod Biomed Online. 2012 Jun;24(6):664-9.

Braam KI, **Overbeek A**, Kaspers GJ, Ronckers CM, Schouten-van Meeteren AY, Van Dulmen-Den Broeder E, Veening MA. Malignant melanoma as a second malignant neoplasm in long-term childhood cancer survivors: A systematic review. Pediatr Blood Cancer. 2012 May;58(5):665-74.

van den Berg MH, **Overbeek A**, van Dulmen-den Broeder E, Lambalk CB. Comment on: Li HW, Wong CY, Yeung WS, Ho PC, Ng EH. Serum anti-Mullerian hormone level is not altered in women using hormonal contraceptives. Contraception. 2012 May;85(5):519-20; author reply 520-1.

Van den Berg MH, **Overbeek A**, van der Pal HP, Versluys AB, Bresters D, Van Leeuwen FE, Lambalk CB, Kaspers GJ, van Dulmen-den Broeder E. Using web-based and paper-based questionnaires for collecting data on fertility issues among female cancer survivors: differences in response characteristics. J Med Internet Res. 2011 Sep 29;13(3):e76.

Overbeek A, van den Berg MH, van Leeuwen FE, Lambalk CB, Kaspers GJL, van Dulmen-den Broeder E. Vruchtbaarheidsproblematiek bij vrouwelijke overlevenden van kinderkanker in Nederland: een pilotstudie. Nederlands Tijdschrift voor Obstetrie en Gynaecologie. 2011 Apr;124:95-101.

van den Berg MH, van Dulmen-den Broeder E, **Overbeek A**, Twisk JW, Schats R, van Leeuwen FE, Kaspers GJ, Lambalk CB. Comparison of ovarian function markers in users of hormonal contraceptives during the hormone-free interval and subsequent natural early follicular phases. Hum Reprod. 2010 Jun;25(6):1520-7.

Kuijper EA, Blankenstein MA, Luttikhof LJ, Roek SJ, **Overbeek A,** Hompes PG, Twisk JW, Lambalk CB. Frequency distribution of polymorphisms in the FSH receptor gene in infertility patients of different ethnicity. Reprod Biomed Online. 2010 May;20(5):588-93.

Overbeek A, Lambalk CB. Pharmacogenomics of ovulation induction: facilitating decisions on who, when and how to treat. Pharmacogenomics. 2009 Sep;10(9):1377-9.

Overbeek A, Kuijper EA, Hendriks ML, Blankenstein MA, Ketel IJ, Twisk JW, Hompes PG, Homburg R, Lambalk CB. Clomiphene citrate resistance in relation to follicle-stimulating hormone receptor Ser680Ser-polymorphism in polycystic ovary syndrome. Hum Reprod. 2009 Aug;24(8):2007-13.

Overbeek A, Lambalk CB. Phenotypic and pharmacogenetic aspects of ovulation induction in WHO II anovulatory women. Gynecol Endocrinol. 2009 Apr;25(4):222-34.

516278-L-bw-Overbeek
Processed on: 19-1-2018

DANKWOORD

En daar zit ik dan, diep in de nacht, tijdens een rustige dienst, mijn dankwoord te schrijven. Tijd om terug te kijken op mijn promotietijd. Het duurt altijd te lang en af en toe waren er momenten dat ik mijn koffiebeker met de slogan van Marja-Liisa echt nodig had: Promoveren is wél leuk! Maar wat heb ik een fijne tijd gehad, wat heb ik veel geleerd en wat hebben veel mensen zich bekommerd om mijn project. Ik ben deze mensen dan ook enorm dankbaar.

Allereerst wil ik alle vrouwen bedanken die hebben meegedaan met het DCOG LATER-VEVO onderzoek. Tevens veel dank voor de huisartsen die hebben meegeholpen om vrouwen voor de controlegroep te werven. En natuurlijk alle specialisten die de moeite hebben genomen de PAK-vragenlijst in te vullen.

Drie promotoren maar liefst hebben mij begeleid bij deze promotie. Gert-Jan, vaak vanaf de zijlijn betrokken, maar daarom des te meer oog voor het grotere geheel. Onze (twee)maandelijkse gesprekjes waren altijd een goed moment om te reflecteren. Ook dank voor je commentaar op mijn stukken. Ik heb je snelle en to-the-point reacties altijd zeer gewaardeerd. Floor, bedankt voor je scherpe blik en kritische commentaar. Ondanks je overvolle agenda lukte het toch altijd om tijd vrij te maken voor DCOG LATER-VEVO. Briljant hoeveel goede ideeën de revue passeerden tijdens onze projectleidersoverleggen. Last but definitely not least: Nils, vanaf 2004 ben ik al onder jouw hoede. Eerst een wetenschappelijke stage, toen protocollendokter en daarna een promotie. Ik bewonder je om je creatieve geest en je manier van 'out of the box' denken. Altijd vol nieuwe ideeën en hypotheses. We hebben inmiddels heel wat samen meegemaakt. Ik denk glimlachend terug aan ons dinertje met Tamar, Iris en Els in Rome bij dat veel te dure restaurant ("de rekening sturen we wel naar Peter"). Of die keer dat ik je bijna over je heen reed omdat jij handmatig, liggend onder je oude Volvo, de auto in de versnelling zette op het parkeerterrein van het NKI. Ik hoop in de toekomst nog veel samen te werken en van je te blijven leren!
Mijn copromotor, Eline. We hebben een hele andere manier van werken, ik volgens stricte planning en klaar-is-klaar, jij last-minute en "zullen we niet toch nog kijken of...". Maar uiteindelijk hebben we wel een prachtig proefschrift afgeleverd, en dat was zonder jou niet gelukt. Hartelijk dank voor al je hulp, kritisch noten en je nauwkeurige correcties.

Mijn andere copromotor, Marleen, die volgens het promotiereglement geen copromotor mag zijn, maar wat mij betreft die eer meer dan verdiend heeft. Marleen, hartelijk dank voor alle tijd en energie die je in het DCOG LATER-VEVO onderzoek hebt gestoken. Jij verdient het doctoraat bijna nog meer dan ik! Ik heb enorm veel van je geleerd. Jij bent waarschijnlijk even blij als ik dat het nu af is! Ik had het niet zonder je gered.

Alle centra die hebben meegewerkt aan het slagen van het DCOG LATER-VEVO onderzoek. In het bijzonder wil ik Wendy van Dorp bedanken. Je inzet voor DCOG LATER-VEVO was enorm en ik heb veel respect voor hoe je alles hebt uitgevoerd.

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 255

Nooit te beroerd om iets uit te zoeken, tot laat in de avond patiënten aan het nabellen. En in de tussentijd nog eerder promoveren dan ik! Heel veel succes met je opleiding tot gynaecoloog. Verder natuurlijk in het Rotterdamse: Marry van den Heuvel en Joop Laven, dank voor jullie inbreng en bijdrage aan het DCOG LATER-VEVO onderzoek. In het UMCG: Gea Huizinga, Wim Tissing, Arnold Simons, Elsbeth Dul, Meike Mutsaers, Tineke Rijken, Edith de Bruin. In het UMCN: Jos Bökkerink, Jacqueline Loonen, Ina Beerendonk, Thea ten Haaf, Ellen van der Vorst, Karin van Mook, Jolieneke Schoonenberg-Pomper. In het AMC: Leontien Kremer, Nino Tonch, Heleen van der Pal, Cor van den Bos, Minke Mud, Elske Sieswerda, Richard Heinen, Ilse Overbeeke. In het LUMC: Dorine Bresters, Marloes Louwerens, Ilse van Gils, Dominique Wanders. In het UMCU: Martine van Engelen, Brigitta Versluys.

Het bureau SKION Later: Cecile Ronckers, Bep Verkerk, Sebastiaan Knijnenburg, Nynke Hollema, Judith Kok, Lieke Feijen, Jop Teepen, Margriet van de Heijden.

Alle co-auteurs wil ik bedanken voor hun kritische blik op alle manuscripten. Jos en Michael: dank voor het scheppen van duidelijkheid in de statistische chaos.

De ene vraag leidt tot een ander, zo werd het PAK project geboren. Toen bleek dat er een soortgelijk idee in het LUMC was, heeft dat een vruchtbare samenwerking opgeleverd. Leoni, Moniek, Anne en Carina, hartelijk dank voor de fijne samenwerking voor het PAK project. Ook hartelijk dank aan het IKNL en SKION voor het beschikbaar stellen van de adresgegevens.

Philips, voor de financiële ondersteuning door middel van het beschikbaar stellen van de echo-apparatuur. Durex, voor het beschikbaar stellen van 30.000 condooms.

Elise en Priyanta, jullie hebben mij zoveel werk uit handen genomen. Die enorme database is mede dankzij jullie nu goud waard! Shabir, dank voor je uren achter de computer om alle eieren te tellen!

Marloes, mijn “opvolger”. Hartelijk dank voor het monnikenwerk dat je verricht heb bij het cleanen van de data. Ik hoop dat we veel artikelen uit deze database samen kunnen gaan schrijven!

Mirte, hartelijk dank voor het maken van de echo’s in mijn afwezigheid.

Ted Korsen, dank voor het afnemen van de bloedmonsters, wanneer nodig.

Lieve Tamar, mijn kamergenootje en congres-buddy. Shoppen, sightseeing, heel veel kletsen en, oh ja, wetenschap. Wat was het altijd gezellig! En zo blijft het, volgend jaar lekker naar de ESHRE in Barcelona?? En bij het volgende gynaecongres weer een kamer delen?

Vanaf april 2012 ben ik in het MCA (nu NWZ) aan de slag gegaan, eerst als ANIOS, daarna als AIOS. Veel wetenschappelijke ervaring, maar klinisch kon ik nog niets. Alle gynaecologen, arts-assistenten, verloskundigen en verpleegkundigen: super bedankt dat jullie mij de kneepjes van het vak wilden leren. Kijk Piet, mijn proefschrift is af. De term 'boekje' doet inderdaad afbreuk aan het bloed, het zweet en de tranen die erin zitten. Nu alweer twee jaar weer terug op het nest in het VUmc. Dank voor de begeleiding bij mijn opleiding en de ruimte voor wetenschappelijke ontwikkeling.

PK 4 X, het is een begrip. Zelden zie je zo'n leuke groep bij elkaar. De lunchdiscussies, de taart-momentjes, de borrels, de sushi, de etentjes. Wat had ik zonder Hester, Gerrit, Susanne, Willemijn, Raphaele, Ilse, Hannemieke, Tamara, Sophie, Rik, Monique, Stephanie, Manita en Sandra gemoeten?

Marc. Grote bergen zoute peren liggen in de zonneschijn, naast een kussen vol met veren en een fles met wijnazijn. Dozen vol met sperziebonen en een klaproos van beton, wie wil in een fietsband wonen, rode biet op het balkon.... Vrijdagmiddag, PK 4X 027. Priceless... Need I say more?

Lieve Katja, bijna vanaf het begin hebben we een kamer gedeeld. Al snel konden we ons hart bij elkaar luchten. Ik heb heel wat stoom kunnen afblazen dankzij jou. Zonder jou was promoveren lang zo leuk niet geweest. Ik bewonder je doorzettingsvermogen en hoop dat je je professionele dromen waar kan gaan maken. Zo fijn dat je mijn paranimf wilt zijn.

Lieve Lisette, mijn paranimf. Al zo lang mijn vriendinnetje, door dik en dun. We delen telkens weer de belangrijkste dingen in ons leven. Met de geboorte van onze kindjes zijn we nog dichter naar elkaar gegroeid. Voltallige promotiecommissie of Bryan in Ahoy: als jij naast me staat, is het sowieso een feestje!

Lieve papa en mama, hartelijk dank voor jullie eeuwige vertrouwen in mijn kunnen. Altijd een luisterend oor, altijd opbeurende woorden en constructieve tips. Ik weet dat jullie je soms zorgen maken of ik niet te veel hooi op mijn vork neem. Ja dus, maar nu is het klaar! Kom maar gauw eten en de nieuwe KitchenAid uitproberen!

Liefste Jef, mijn held. Je hebt me er doorheen gesleept, vooral de laatste loodjes toen ik de combi van opleiding, promotie, moeder en partner niet altijd meer zag zitten. Altijd klaar voor een knuffel of een kus, altijd het eten klaar als ik weer eens laat thuis was of door moest werken. Door dik en dun. Je bent mijn alles, ik wil nooit meer zonder je. Ik hou van je! Nu kunnen we eindelijk de bruiloft gaan plannen!

Jasper en Ruben, mijn schatjes, jullie tonen me elke dag wat écht belangrijk is in het leven.

CURRICULUM VITAE

Annelies Overbeek was born April 15th 1983, in Alkmaar, The Netherlands. She obtained her medical degree at the VU University Faculty of Medicine in Amsterdam in 2006. For two years she worked at the IVF department of the Reproductive Medicine Unit of the VU University Medical Center. In 2008 she started as a PhD student with the research described in this dissertation at the department of pediatric oncology at the VU University Medical Center. The project was supervised by prof. dr. Gertjan Kaspers, prof. dr. Flora van Leeuwen and prof. dr. Nils Lambalk. During this period, she obtained a master's degree in Epidemiology. From January 2012 to October 2012, she worked as a resident (ANIOS) in Obstetrics and Gynecology in the Medisch Centrum Alkmaar in Alkmaar. She started her residency training in obstetrics and gynecology in January 2013 at the Medisch Centrum Alkmaar (supervised by dr. Joke Bais) and in January 2016 at the VU University Medical Center (supervised by prof. dr. Hanneke de Vries). She is engaged to Jeffrey van der Steen and they have two children, Jasper and Ruben.

516278-L-bw-Overbeek
Processed on: 19-1-2018
PDF page: 259

R

References

REFERENCES

1. Wallace WH, Kelsey TW. Human ovarian reserve from conception to the menopause. PloS one 2010;5(1):e8772 doi: 10.1371/journal.pone.0008772[published Online First: Epub Date]|.
2. Henderson KD, Bernstein L, Henderson B, et al. Predictors of the timing of natural menopause in the Multiethnic Cohort Study. Am J Epidemiol 2008;167(11):1287-94 doi: 10.1093/aje/kwn046[published Online First: Epub Date]|.
3. Lambalk CB, van DJ, de Koning CH, et al. Testing ovarian reserve to predict age at menopause. Maturitas 2009;63(4):280-91
4. Knight PG, Glister C. TGF-beta superfamily members and ovarian follicle development. Reproduction 2006;132(2):191-206 doi: 10.1530/rep.1.01074[published Online First: Epub Date]|.
5. Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. Endocr Rev 1996;17(2):121-55 doi: 10.1210/edrv-17-2-121[published Online First: Epub Date]|.
6. Durlinger AL, Gruijters MJ, Kramer P, et al. Anti-Mullerian hormone inhibits initiation of primordial follicle growth in the mouse ovary. Endocrinology 2002;143(3):1076-84 doi: 10.1210/endo.143.3.8691[published Online First: Epub Date]|.
7. Casper RF, Mitwally MF. Review: aromatase inhibitors for ovulation induction. J Clin Endocrinol Metab 2006;91(3):760-71 doi: 10.1210/jc.2005-1923[published Online First: Epub Date]|.
8. Verhoeven MOL, C.B. Ovarian Physiology. In: Belfiore AL, D., ed. Principles of Endocrinology and Hormone Action: Springer International Publishing, 2016:1-22.
9. McGee EA, Hsueh AJ. Initial and cyclic recruitment of ovarian follicles. Endocr Rev 2000;21(2):200-14 doi: 10.1210/edrv.21.2.0394[published Online First: Epub Date]|.
10. Lambalk CB. Anti-Müllerian hormone, the holy grail for fertility counselling in the general population? Human Reproduction 2015;30(10):2257-58 doi: 10.1093/humrep/dev199[published Online First: Epub Date]|.
11. Tinkanen H, Blauer M, Laippala P, et al. Correlation between serum inhibin B and other indicators of the ovarian function. Eur J Obstet Gynecol Reprod Biol 2001;94(1):109-13
12. Danforth DR, Arbogast LK, Mroueh J, et al. Dimeric inhibin: a direct marker of ovarian aging. Fertil Steril 1998;70(1):119-23
13. Burger HG, Hale GE, Robertson DM, et al. A review of hormonal changes during the menopausal transition: focus on findings from the Melbourne Women's Midlife Health Project. Human reproduction update 2007;13(6):559-65 doi: 10.1093/humupd/dmm020[published Online First: Epub Date]|.
14. Durlinger AL, Gruijters MJ, Kramer P, et al. Anti-Mullerian hormone attenuates the effects of FSH on follicle development in the mouse ovary. Endocrinology 2001;142(11):4891-99
15. Durlinger AL, Kramer P, Karels B, et al. Control of primordial follicle recruitment by anti-Mullerian hormone in the mouse ovary. Endocrinology 1999;140(12):5789-96
16. Weenen C, Laven JS, Von Bergh AR, et al. Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. MolHumReprod 2004;10(2):77-83
17. Hamre H, Kiserud CE, Ruud E, et al. Gonadal function and parenthood 20 years after treatment for childhood lymphoma: a cross-sectional study. Pediatr Blood Cancer 2012;59(2):271-77
18. Rendtorff R, Hohmann C, Reinmuth S, et al. Hormone and Sperm analyses after chemo- and radiotherapy in Childhood and Adolescence. Klin Padiatr 2010;222:145-49
19. van Waas M, Neggers SJ, Te Winkel ML, et al. Endocrine late sequelae in long-term survivors of childhood non-Hodgkin lymphoma. Ann Oncol 2012;23(6):1626-32

20. Rustamov O, Smith A, Roberts SA, et al. The measurement of anti-Mullerian hormone: a critical appraisal. J Clin Endocrinol Metab 2014;99(3):723-32 doi: 10.1210/jc.2013-3476[published Online First: Epub Date]|.
21. La Marca A, Stabile G, Artenisio AC, et al. Serum anti-Mullerian hormone throughout the human menstrual cycle. HumReprod 2006;21(12):3103-07
22. Tsepelidis S, Devreker F, Demeestere I, et al. Stable serum levels of anti-Mullerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. HumReprod 2007;22(7):1837-40
23. Cook, C.L, Siow, et al. Serum mullerian-inhibiting substance levels during normal menstrualcycles. Fertil Steril 2000;73:859-61
24. Gnoth C, Roos J, Broomhead D, et al. Antimullerian hormone levels and numbers and sizes of antral follicles in regularly menstruating women of reproductive age referenced to true ovulation day. Fertility and sterility 2015;104(6):1535-43 e1-4 doi: 10.1016/j.fertnstert.2015.08.027[published Online First: Epub Date]|.
25. Hadlow N, Longhurst K, McClements A, et al. Variation in antimüllerian hormone concentration during the menstrual cycle may change the clinical classification of the ovarian response. Fertility and Sterility 2012;99(6):1791-97
26. Overbeek A, Broekmans FJ, Hehenkamp WJ, et al. Intra-cycle fluctuations of anti-Müllerian hormone in normal women with a regular cycle: a re-analysis. Reproductive biomedicine online 2012;24(6):664-69
27. Streuli, I, Fraisse, et al. Clinical uses of anti-Mullerian hormone assays:pitfalls and promises. Fertil Steril 2009;91:226-30
28. Wunder, D.M, Bersinger, et al. Statistically significant changes of antimullerianhormone and inhibin levels during the physiologic menstrualcycle in reproductive age women. Fertil Steril 2008;89:927-33
29. Hansen KR, Hodnett GM, Knowlton N, et al. Correlation of ovarian reserve tests with histologically determined primordial follicle number. Fertility and Sterility 2011;95(1):170-75 doi: 10.1016/j.fertnstert.2010.04.006[published Online First: Epub Date]|.
30. Hagen CP, Vestergaard S, Juul A, et al. Low concentration of circulating antimullerian hormone is not predictive of reduced fecundability in young healthy women: a prospective cohort study. Fertility and sterility 2012;98(6):1602-8 e2 doi: 10.1016/j.fertnstert.2012.08.008[published Online First: Epub Date]|.
31. Depmann M, Eijkemans MJ, Broer SL, et al. Does anti-Mullerian hormone predict menopause in the general population? Results of a prospective ongoing cohort study. Hum Reprod 2016;31(7):1579-87 doi: 10.1093/humrep/dew112[published Online First: Epub Date]|.
32. Basisregistratie S. SKION Nederland, 2017.
33. Smith M, Hare ML. An overview of progress in childhood cancer survival. Journal of pediatric oncology nursing : official journal of the Association of Pediatric Oncology Nurses 2004;21(3):160-4
34. Stiller CA. Population based survival rates for childhood cancer in Britain, 1980-91. BMJ 1994;309(6969):1612-6
35. Terracini B, Coebergh JW, Gatta G, et al. Childhood cancer survival in Europe: an overview. Eur J Cancer 2001;37(6):810-6
36. Geenen MM, Cardous-Ubbink MC, Kremer LC, et al. Medical assessment of adverse health outcomes in long-term survivors of childhood cancer. JAMA 2007;297(24):2705-15
37. Oeffinger KC, Mertens AC, Sklar CA, et al. Chronic health conditions in adult survivors of childhood cancer. NEnglJMed 2006;355(15):1572-82
38. Turcotte LM, Liu Q, Yasui Y, et al. Temporal Trends in Treatment and Subsequent Neoplasm Risk Among 5-Year Survivors of Childhood Cancer, 1970-2015. JAMA 2017;317(8):814-24 doi: 10.1001/jama.2017.0693[published Online First: Epub Date]|.

39. Bath LEW. Late effects of the treatment of childhood cancer on the female reproductive system and the potential for fertility preservation. BJOG : an international journal of obstetrics and gynaecology 2002(2):107-14
40. Whitehead E, Shalet SM, Blackledge G, et al. The effect of combination chemotherapy on ovarian function in women treated for Hodgkin's disease. Cancer 1983;52(6):988-93
41. Chow EJ, Liu W, Srivastava K, et al. Differential effects of radiotherapy on growth and endocrine function among acute leukemia survivors: a childhood cancer survivor study report. Pediatr Blood Cancer 2013;60(1):110-5 doi: 10.1002/pbc.24198[published Online First: Epub Date]|.
42. Green DM, Whitton JA, Stovall M, et al. Pregnancy outcome of female survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. Am J Obstet Gynecol 2002;187:1070 - 80
43. Thomson AB, Critchley HO, Kelnar CJ, et al. Late reproductive sequelae following treatment of childhood cancer and options for fertility preservation. Best practice & research Clinical endocrinology & metabolism 2002;16(2):311-34 doi: 10.1053/beem.2002.0200[published Online First: Epub Date]|.
44. Nielsen SN, Andersen AN, Schmidt KT, et al. A 10-year follow up of reproductive function in women treated for childhood cancer. ReprodBiomedOnline 2013;27(2):192-200
45. Chemaitilly W, Mertens AC, Mitby P, et al. Acute ovarian failure in the childhood cancer survivor study. J Clin EndocrinolMetab 2006;91(5):1723-28
46. Byrne J, Fears TR, Gail MH, et al. Early menopause in long-term survivors of cancer during adolescence. AmJ ObstetGynecol 1992;166(0002-9378 (Print), 0002-9378 (Linking), 3):788-93
47. Thomas-Teinturier C, El Fayech C, Oberlin O, et al. Age at menopause and its influencing factors in a cohort of survivors of childhood cancer: earlier but rarely premature. Hum Reprod 2013;28(2):488-95 doi: 10.1093/humrep/des391[published Online First: Epub Date]|.
48. Chemaitilly W, Li Z, Krasin MJ, et al. Premature Ovarian Insufficiency in Childhood Cancer Survivors: A Report from the St. Jude Lifetime Cohort. J Clin Endocrinol Metab 2017 doi: 10.1210/jc.2016-3723[published Online First: Epub Date]|.
49. Wallace WH, Thomson AB, Saran F, et al. Predicting age of ovarian failure after radiation to a field that includes the ovaries. Int J Radiat Oncol Biol Phys 2005;62:738 - 44
50. Meirow D. Reproduction post-chemotherapy in young cancer patients. MolCell Endocrinol 2000;169(1-2):123-31
51. Meirow D, Biederman H, Anderson RA, et al. Toxicity of chemotherapy and radiation on female reproduction. ClinObstetGynecol 2010;53(4):727-39
52. Ben-Aharon I, Shalgi R. What lies behind chemotherapy-induced ovarian toxicity? Reproduction 2012;144:153-63
53. Sklar CA, Mertens AC, Mitby P, et al. Premature menopause in survivors of childhood cancer: a report from the childhood cancer survivor study. J NatlCancer Inst 2006;98(13):890-96
54. Charpentier AM, Chong AL, Gingras-Hill G, et al. Anti-Mullerian hormone screening to assess ovarian reserve among female survivors of childhood cancer. Journal of cancer survivorship : research and practice 2014;8(4):548-54 doi: 10.1007/s11764-014-0364-4[published Online First: Epub Date]|.
55. Gracia CR, Sammel MD, Freeman E, et al. Impact of cancer therapies on ovarian reserve. Fertility and Sterility 2012;97(1):134-40
56. Lie Fong S, Lugtenburg PJ, Schipper I, et al. Anti-mullerian hormone as a marker of ovarian function in women after chemotherapy and radiotherapy for haematological malignancies. Hum Reprod 2008;23:674 - 78

57. Thomas-Teinturier C, Allodji RS, Svetlova E, et al. Ovarian reserve after treatment with alkylating agents during childhood. Hum Reprod 2015;30(6):1437-46 doi: 10.1093/humrep/dev060[published Online First: Epub Date]|.
58. van Dorp W, Mulder RL, Kremer LC, et al. Recommendations for Premature Ovarian Insufficiency Surveillance for Female Survivors of Childhood, Adolescent, and Young Adult Cancer: A Report From the International Late Effects of Childhood Cancer Guideline Harmonization Group in Collaboration With the PanCareSurFup Consortium. J Clin Oncol 2016;34(28):3440-50 doi: 10.1200/JCO.2015.64.3288[published Online First: Epub Date]|.
59. Bath LE, Wallace WHB, Shaw MP, et al. Depletion of ovarian reserve in young women after treatment for cancer in childhood: Detection by anti-Mullerian hormone, inhibin B and ovarian ultrasound. Hum Reprod 2003;18:2368 - 74
60. Lantinga GM, Simons AH, Kamps WA, et al. Imminent ovarian failure in childhood cancer survivors. Eur J Cancer 2006;42:1415 - 20
61. Larsen EC, Muller J, Rechnitzer C, et al. Diminished ovarian reserve in female childhood cancer survivors with regular menstrual cycles and basal FSH <10 IU/l. Hum Reprod 2003;18:417 - 22
62. Byrne J, Mulvihill JJ, Myers MH, et al. Effects of treatment on fertility in long-term survivors of childhood or adolescent cancer. NEnglJ Med 1987;317(0028-4793 (Print), 0028-4793 (Linking), 21):1315-21
63. Chiarelli AM, Marrett LD, Darlington G. Early menopause and infertility in females after treatment for childhood cancer diagnosed in 1964-1988 in Ontario, Canada. Am J Epidemiol 1999;150:245 - 54
64. Green DM, Sklar CA, Boice JD, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: results from the Childhood Cancer Survivor Study. J Clin Oncol 2009;27:2374 - 81
65. Reinmuth S, Liebeskind AK, Wickmann L, et al. Having children after surviving cancer in childhood or adolescence - results of a Berlin survey. Klin Padiatr 2008;220:159 - 65
66. Meirow D, Dor J, Kaufman B, et al. Cortical fibrosis and blood-vessels damage in human ovaries exposed to chemotherapy. Potential mechanisms of ovarian injury. Hum Reprod 2007;22(6):1626-33 doi: 10.1093/humrep/dem027[published Online First: Epub Date]|.
67. Lee SJ, Schover LR, Partridge AH, et al. American Society of Clinical Oncology recommendations on fertility preservation in cancer patients. J Clin Oncol 2006;24(18):2917-31 doi: 10.1200/JCO.2006.06.5888[published Online First: Epub Date]|.
68. Lambertini M, Del Mastro L, Pescio MC, et al. Cancer and fertility preservation: international recommendations from an expert meeting. BMC Med 2016;14:1 doi: 10.1186/s12916-015-0545-7[published Online First: Epub Date]|.
69. Nelson LM. Clinical practice. Primary ovarian insufficiency. New England Journal of Medicine 2009;360:606-14
70. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet 2002;359:341-45
71. Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. BMC Medical Research Methodology 2007;7:32-32
72. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from wwwcochrane-handbookorg.
73. Laverdiere C, Liu Q, Yasui Y, et al. Long-term outcomes in survivors of neuroblastoma: A report from the childhood cancer survivor study. Journal of the National Cancer Institute 2009;101:1131-40
74. Andrieu JM, Ochoa-Molina ME. Menstrual cycle, pregnancies and offspring before and after MOPP therapy for Hodgkin's disease. Cancer 1983;52(3):435-8

75. Behringer K, Breuer K, Reineke T, et al. Secondary amenorrhea after Hodgkin's lymphoma is influenced by age at treatment, stage of disease, chemotherapy regimen, and the use of oral contraceptives during therapy: a report from the German Hodgkin's Lymphoma Study Group. J Clin Oncol 2005;23(30):7555-64
76. Behringer K, Thielen I, Mueller H, et al. Fertlity and gonadal function in female survivors after treatment of early unfavorable Hodgkin lymphoma (HL) within the German Hodgkin Study Group HD14 trial. Annals of Oncology 2012;23:1818-25
77. Bresters D, Emons JA, Nuri N, et al. Ovarian insufficiency and pubertal development after hematopoietic stem cell transplantation in childhood. Pediatr Blood Cancer 2014;61(11):2048-53 doi: 10.1002/pbc.25162[published Online First: Epub Date]|.
78. Brignardello E, Felicetti F, Castiglione A et al. Endocrine health conditions in adult survivors of childhood cancer: the need for specialized adult-focused follow-up clinics. Eur J Endocrinol 2013;168:465-72
79. Chapman RM, Sutcliffe SB, Malpas JS. Cytotoxic-induced ovarian failure in women with Hodgkin's disease. I. Hormone function. JAMA 1979;242(17):1877-81
80. De Bruin ML, Huisbrink J, Hauptmann M, et al. Treatment-related risk factors for premature menopause following Hodgkin lymphoma. Blood 2008;111(1):101-08
81. Even-Or E, Ben-Haroush A, Yahel A, et al. Fertility After Treatment With High Dose Melphalan in Women With Acute Myelogenous Leukemia. Pediatr Blood Cancer 2015 doi: 10.1002/pbc.25765[published Online First: Epub Date]|.
82. Falorio S, Biasoli I, Luminari S et al. Risk factors for impaired gonadal function in female Hodgkin lymphoma survivors: final analysis of a retrospective multicenter joint study from Italian and Brazilian Institutions. Hematol Oncol 2013;31(332-338)
83. Giuseppe L, Attilio G, Edoardo DN, et al. Ovarian function after cancer treatment in young women affected by Hodgkin disease (HD). Hematology 2007;12(2):141-47
84. Haukvik UKH, Dieset I, Bjøro T, et al. Treatment-related premature ovarian failure as a long-term complication after Hodgkin's lymphoma. Annals of Oncology 2006;17:1428-33
85. Katz D, Hamburger T, Divinsky L, et al. Fertility and pregnancy in patients below the age of 35 following chemotherapy for breast cancer. Cancer Research 2009;Conference: 31st Annual San Antonio Breast Cancer Symposium San Antonio, TX United States. Conference Start: 20081210 Conference End: 20081214 Sponsor: UT Health Science Center San Antonio School of Medicine, American Association for Cancer Research, Baylor College of Medicine. Conference Publication:
86. King DJ, Ratcliffe MA, Dawson AA, et al. Fertility in young men and women after treatment for lymphoma: a study of a population. Journal of clinical pathology 1985;38(11):1247-51
87. Klemp JR, Smith AK, Ranallo L, et al. Baseline characteristics of women initiating follow-up care in a newly developed breast cancer survivorship center. Cancer Research 2009;Conference: 31st Annual San Antonio Breast Cancer Symposium San Antonio, TX United States. Conference Start: 20081210 Conference End: 20081214 Sponsor: UT Health Science Center San Antonio School of Medicine, American Association for Cancer Research, Baylor College of Medicine. Conference Publication:
88. Koyama H, Wada T, Nishizawa Y, et al. Cyclophosphamide-induced ovarian failure and its therapeutic significance in patients with breast cancer. Cancer 1977;39(4):1403-9
89. Krawczuk-Rybak M, Leszczynska E, Poznanska M, et al. Anti-mullerian hormone as a sensitive marker of ovarian function in young cancer survivors. International journal of endocrinology 2013;2013:125080 doi: 10.1155/2013/125080[published Online First: Epub Date]|.
90. Krawczuk-Rybak M, Leszczynska E, Poznanska M, et al. The progressive reduction in the ovarian reserve in young women after anticancer treatment. Horm Metab Res 2013;45:813-19

91. Martin I, Valls C, Pavia C, et al. Gonadal function and puberty assessment in a cohort of Spanish pediatric survivors of childhood cancer. Endocrinologist 2009;19:133-41
92. McCune JS, Games DM, Espirito JL. Assessment of ovarian failure and osteoporosis in premenopausal breast cancer survivors. J Oncol PharmPract 2005;11(2):37-43
93. Minisini AM, Menis J, Valent F, et al. Determinants of recovery from amenorrhea in premenopausal breast cancer patients receiving adjuvant chemotherapy in the taxane era. Anticancer Drugs 2009;20(6):503-07
94. Miyoshi Y, Ohta H, Hashii Y, et al. Endocrinological analysis of 122 Japanese childhood cancer survivors in a single hospital. EndocrJ 2008;55(6):1055-63
95. Muller HL, Klinkhammer-Schalke M, Seelbach-Gobel B, et al. Gonadal function of young adults after therapy of malignancies during childhood or adolescence. EurJ Pediatr 1996;155(9):763-69
96. Papadakis V, Vlachopapadopoulou E, Van Syckle K, et al. Gonadal function in young patients successfully treated for Hodgkin disease. MedPediatrOncol 1999;32(5):366-72
97. Raciborska A, Bilska K, Filipp E, et al. Ovarian function in female survivors after multimodal Ewing sarcoma therapy. Pediatr Blood Cancer 2014 doi: 10.1002/pbc.25304[published Online First: Epub Date]|.
98. Salooja N, Chatterjee R, McMillan AK, et al. Successful pregnancies in women following single autotransplant for acute myeloid leukemia with a chemotherapy ablation protocol. Bone Marrow Transplant 1994;13(4):431-35
99. Schilsky RL, Sherins RJ, Hubbard SM, et al. Long-term follow up of ovarian function in women treated with MOPP chemotherapy for Hodgkin's disease. The American journal of medicine 1981;71(4):552-6
100. Stillman RJ, Schinfeld JS, Schiff I, et al. Ovarian failure in long-term survivors of childhood malignancy. AmJ ObstetGynecol 1981;139(0002-9378 (Print), 0002-9378 (Linking), 1):62-66
101. Swerdlow AJ, Cooke R, Bates A, et al. Risk of premature menopause after treatment for Hodgkin's lymphoma. J Natl Cancer Inst 2014;106(9) doi: 10.1093/jnci/dju207[published Online First: Epub Date]|.
102. Teinturier C, Hartmann O, Valteau-Couanet D, et al. Ovarian function after autologous bone marrow transplantation in childhood: high-dose busulfan is a major cause of ovarian failure. Bone Marrow Transplant 1998;22(10):989-94
103. van der Kaaij MA, Heutte N, Meijnders P, et al. Premature ovarian failure and fertility in long-term survivors of Hodgkin's lymphoma: a European Organisation for Research and Treatment of Cancer Lymphoma Group and Groupe d'Etude des Lymphomes de l'Adulte Cohort Study. J Clin Oncol 2012;30(3):291-99
104. Vatanen A, Wilhelmsson M, Borgstrom B et al. Ovarian function after allogeneic hematopoietic stem cell transplantation in childhood and adolescence. Eur J Endocrinol 2013;170:211-18
105. Waxman JH, Terry YA, Wrigley PF, et al. Gonadal function in Hodgkin's disease: long-term follow-up of chemotherapy. British medical journal 1982;285(6355):1612-3
106. Wikstrom AM, Hovi L, Dunkel L, et al. Restoration of ovarian function after chemotherapy for osteosarcoma. ArchDisChild 2003;88(1468-2044 (Electronic), 0003-9888 (Linking), 5):428-31
107. Zaletel LZ, Bratanic N, Jereb B. Gonadal function in patients treated for Hodgkin's disease in childhood. Radiol Oncol 2010;44(3):187-93
108. Afifi Z, Shaw PJ, Clavano-Harding A, et al. Growth and endocrine function in children with acute myeloid leukaemia after bone marrow transplantation using busulfan/cyclophosphamide. Bone Marrow Transplantation 2000;25:1087-92
109. Green DM, Nolan VG, Goodman PJ, et al. The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: a report from the Childhood Cancer Survivor Study. Pediatr Blood Cancer 2014;61(1):53-67 doi: 10.1002/pbc.24679[published Online First: Epub Date]|.

110. Hamre H, Kiserud CE, Ruud E, et al. Gonadal function and parenthood 20 years after treatment for childhood lymphoma: a cross-sectional study. Pediatric blood & cancer 2012;59(1545-5017 (Electronic), 1545-5009 (Linking), 2):271-77
111. Van den Berg MH, van Dulmen-den BE, Overbeek A, et al. Comparison of ovarian function markers in users of hormonal contraceptives during the hormone-free interval and subsequent natural early follicular phases. HumReprod 2010
112. Letourneau JM, Ebbel EE, Katz PP, et al. Acute ovarian failure underestimates age-specific reproductive impairment for young women undergoing chemotherapy for cancer. Cancer 2012;118:1933-39
113. Loren AW, Mangu PB, Beck LN, et al. Fertility preservation for patients with cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol 2013;31(19):2500-10 doi: 10.1200/JCO.2013.49.2678[published Online First: Epub Date]|.
114. Chemaitilly W, Mertens AC, Mitby P, et al. Acute ovarian failure in the childhood cancer survivor study. J Clin Endocrinol Metab 2006;91:1723 - 28
115. Lie FS, van den Heuvel-Eibrink MM, Eijkemans MJ, et al. Pregnancy outcome in female childhood cancer survivors. Hum Reprod 2010;25:1206 - 12
116. Madanat-Harjuoja LM, Malila N, Lahteenmaki PM, et al. Preterm delivery among female survivors of childhood, adolescent and young adulthood cancer. Int J Cancer 2010;127:1669 - 79
117. Reulen RC, Zeegers MP, Wallace WH, et al. Pregnancy outcomes among adult survivors of childhood cancer in the British Childhood Cancer Survivor Study. Cancer Epidemiol Biomarkers Prev 2009;18:2239 - 47
118. Signorello LB, Mulvihill JJ, Green DM, et al. Stillbirth and neonatal death in relation to radiation exposure before conception: a retrospective cohort study. Lancet 2010;376:624 - 30
119. Sklar CA, Mertens AC, Mitby P, et al. Premature menopause in survivors of childhood cancer: a report from the childhood cancer survivor study. J Natl Cancer Inst 2006;98:890 - 96
120. van Beek RD, van den Heuvel-Eibrink MM, Laven JS, et al. Anti-Mullerian hormone is a sensitive serum marker for gonadal function in women treated for Hodgkin's lymphoma during childhood. J Clin Endocrinol Metab 2007;92:3869 - 74
121. van Dorp W, van Beek RD, Laven JS, et al. Long-term endocrine side effects of childhood Hodgkin's lymphoma treatment: a review. Hum Reprod Update 2012;18:12 - 28
122. Armstrong GT, Whitton JA, Gajjar A, et al. Abnormal timing of menarche in survivors of central nervous system tumors: A report from the Childhood Cancer Survivor Study. Cancer 2009;115:2562 - 70
123. Byrne J, Fears TR, Gail MH, et al. Early menopause in long-term survivors of cancer during adolescence. Am J Obstet Gynecol 1992;166:788 - 93
124. Green DM, Kawashima T, Stovall M, et al. Fertility of female survivors of childhood cancer: a report from the childhood cancer survivor study. J Clin Oncol 2009;27:2677 - 85
125. Magelssen H, Melve KK, Skjaerven R, et al. Parenthood probability and pregnancy outcome in patients with a cancer diagnosis during adolescence and young adulthood. Hum Reprod 2008;23:178 - 86
126. Mueller BA, Chow EJ, Kamineni A, et al. Pregnancy outcomes in female childhood and adolescent cancer survivors: a linked cancer-birth registry analysis. Arch Pediatr Adolesc Med 2009;163:879 - 86
127. Sudour H, Chastagner P, Claude L, et al. Fertility and pregnancy outcome after abdominal irradiation that included or excluded the pelvis in childhood tumor survivors. Int J Radiat Oncol Biol Phys 2010;76:867 - 73
128. Winther JF, Boice JD, Svendsen AL, et al. Spontaneous abortion in a Danish population-based cohort of childhood cancer survivors. JClinOncol 2008;26:4340 - 46
129. Bath LE, Critchley HO, Chambers SE, et al. Ovarian and uterine characteristics after total body irradiation in childhood and adolescence: response to sex steroid replacement. British journal of obstetrics and gynaecology 1999;106:1265 - 72

130. Larsen EC, Muller J, Schmiegelow K, et al. Reduced ovarian function in long-term survivors of radiation- and chemotherapy-treated childhood cancer. J Clin Endocrinol Metab 2003;88:5307 - 14
131. de Boer EJ, Den Tonkelaar I, Burger CW, et al. Validity of self-reported causes of subfertility. Am J Epidemiol 2005;161:978 - 86
132. van Leeuwen FE, Klip H, Mooij TM, et al. Risk of borderline and invasive ovarian tumours after ovarian stimulation for in vitro fertilization in a large Dutch cohort. Hum Reprod 2011;26:3456 - 65
133. Hoonakker P, Carayo P. Questionnaire Survey Nonresponse: A Comparison of Postal Mail and Internet Surveys. Int J Hum Comput Interact 2009;25:348 - 73
134. Pealer LN, Weiler RM, Pigg RM, et al. The feasibility of a web-based surveillance system to collect health risk behavior data from college students. Health Educ Behav 2001;28:547 - 59
135. Vereecken CA, Maes L. Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. J Adolesc Health 2006;38:426 - 32
136. van den Berg MH, Overbeek A, van der Pal HJ, et al. Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics. J Med Internet Res 2011;13:e76
137. Coolman M, de Groot CJ, Jaddoe VW, et al. Medical record validation of maternally reported history of preeclampsia. J Clin Epidemiol 2010;63:932 - 37
138. Olson JE, Shu XO, Ross JA, et al. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. Am J Epidemiol 1997;145:58 - 67
139. Rice F, Lewis A, Harold G, et al. Agreement between maternal report and antenatal records for a range of pre and peri-natal factors: the influence of maternal and child characteristics. Early Hum Dev 2007;83:497 - 504
140. Overbeek A, van den Berg MH, Hukkelhoven C, et al. Validity of self-reported data on pregnancies: a comparison of self-report with data from a nationwide population-based registry. Accepted
141. Jayaprakasan K, Campbell BK, Clewes JS, et al. Three-dimensional ultrasound improves the interobserver reliability of antral follicle counts and facilitates increased clinical work flow. Ultrasound Obstet Gynecol 2008;31:439 - 44
142. Jayaprakasan K, Walker KF, Clewes JS, et al. The interobserver reliability of off-line antral follicle counts made from stored three-dimensional ultrasound data: a comparative study of different measurement techniques. Ultrasound Obstet Gynecol 2007;29:335 - 41
143. Merce LT, Gomez B, Engels V, et al. Intraobserver and interobserver reproducibility of ovarian volume, antral follicle count, and vascularity indices obtained with transvaginal 3-dimensional ultrasonography, power Doppler angiography, and the virtual organ computer-aided analysis imaging program. J Ultrasound Med 2005;24:1279 - 87
144. Scheffer GJ, Broekmans FJ, Bancsi LF, et al. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound Obstet Gynecol 2002;20:270 - 75
145. Ness KK, Leisenring W, Goodman P, et al. Assessment of selection bias in clinic-based populations of childhood cancer survivors: a report from the childhood cancer survivor study. Pediatr Blood Cancer 2009;52:379 - 86
146. Gatta G, Zigon G, Capocaccia R, et al. Survival of European children and young adults with cancer diagnosed 1995-2002. EurJCancer 2009;45(6):992-1005
147. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2008. Secondary SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2008 2011.

148. Bhatia S, Landier W. Evaluating survivors of pediatric cancer. Cancer J 2005;11(4):340-54
149. Oeffinger KC, van Leeuwen FE, Hodgson DC. Methods to assess adverse health-related outcomes in cancer survivors. Cancer EpidemiolBiomarkers Prev 2011;20(10):2022-34
150. Robison LL. Methodologic issues in the study of second malignant neoplasms and pregnancy outcomes. MedPediatrOncolSuppl 1996;1:41-44
151. Hawkins MM, Lancashire ER, Winter DL, et al. The British Childhood Cancer Survivor Study: Objectives, methods, population structure, response rates and initial descriptive information. PediatrBlood Cancer 2008;50(5):1018-25
152. Hudson MM, Ness KK, Nolan VG, et al. Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. PediatrBlood Cancer 2011;56(5):825-36
153. McBride ML, Rogers PC, Sheps SB, et al. Childhood, adolescent, and young adult cancer survivors research program of British Columbia: objectives, study design, and cohort characteristics. PediatrBlood Cancer 2010;55(2):324-30
154. Robison LL, Armstrong GT, Boice JD, et al. The Childhood Cancer Survivor Study: a National Cancer Institute-supported resource for outcome and intervention research. JClinOncol 2009;27(14):2308-18
155. Robison LL, Mertens AC, Boice JD, et al. Study design and cohort characteristics of the Childhood Cancer Survivor Study: a multi-institutional collaborative project. MedPediatrOncol 2002;38(4):229-39
156. Shaw AK, Morrison HI, Speechley KN, et al. The late effects study: design and subject representativeness of a Canadian, multi-centre study of late effects of childhood cancer. ChronicDisCan 2004;25(3-4):119-26
157. Hawkins MM, Robison LL. Importance of clinical and epidemiological research in defining the long-term clinical care of pediatric cancer survivors. PediatrBlood Cancer 2006;46(2):174-78
158. Leisenring WM, Mertens AC, Armstrong GT, et al. Pediatric cancer survivorship research: experience of the Childhood Cancer Survivor Study. JClinOncol 2009;27(14):2319-27
159. Lund LW, Schmiegelow K, Rechnitzer C, et al. A systematic review of studies on psychosocial late effects of childhood cancer: structures of society and methodological pitfalls may challenge the conclusions. PediatrBlood Cancer 2011;56(4):532-43
160. Murphy SB. Research involving long-term survivors of childhood and adolescent cancer: issues impacting design and conduct of clinical trials. CurrProblCancer 2003;27(4):225-35
161. Ness KK, Leisenring W, Goodman P, et al. Assessment of selection bias in clinic-based populations of childhood cancer survivors: a report from the childhood cancer survivor study. PediatrBlood Cancer 2009;52(3):379-86
162. Overbeek A, van den Berg MH, Kremer LC, et al. A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges. BMCCancer 2012;12(1):363
163. Buchbinder D, Casillas J, Krull KR, et al. Psychological outcomes of siblings of cancer survivors: a report from the Childhood Cancer Survivor Study. Psychooncology 2011;20(12):1259-68
164. Cuttini M, Da FM, Haupt R, et al. Survivors of childhood cancer: using siblings as a control group. Pediatrics 2003;112(6 Pt 1):1454-55
165. Goldstein J, Lutz W, Testa MR. The emergence of Sub-Replacement Family Size Ideals in Europe. Population Research and Policy Review 2003;22(5-6):479-96
166. Eilegård A, Steineck G, Nyberg T, et al. Psychological health in siblings who lost a brother or sister to cancer 2 to 9 years earlier. Psycho-Oncology 2013;22(3):683-91 doi: 10.1002/pon.3053[published Online First: Epub Date]|.

167. Prchal A, Landolt MA. How siblings of pediatric cancer patients experience the first time after diagnosis: a qualitative study. Cancer nursing 2012;35(2):133-40 doi: 10.1097/NCC.0b013e31821e0c59[published Online First: Epub Date]|.
168. Scott DA, Valery PC, Boyle FM, et al. Does research into sensitive areas do harm? Experiences of research participation after a child's diagnosis with Ewing's sarcoma. The Medical journal of Australia 2002;177(9):507-10
169. Ward HJ, Cousens SN, Smith-Bathgate B, et al. Obstacles to conducting epidemiological research in the UK general population. BMJ 2004;329(7460):277-79
170. Boivin J, Bunting L, Collins JA, et al. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Human Reproduction 2007;22(6):1506-12 doi: 10.1093/humrep/dem046[published Online First: Epub Date]|.
171. Shucksmitha M, Camerona S, ridewa T, et al. Urban-Rural Differences in Quality of Life across the European Union. Regional Studies 2009;43(10):1275-89
172. Correia S, Rodrigues T, Barros H. Socioeconomic variations in female fertility impairment: a study in a cohort of Portuguese mothers. BMJ Open 2014;4(1):e003985
173. Alderfer MA, Labay LE, Kazak AE. Brief report: does posttraumatic stress apply to siblings of childhood cancer survivors? JPediatrPsychol 2003;28(4):281-86
174. Houtzager BA, Grootenhuis MA, Caron HN, et al. Quality of life and psychological adaptation in siblings of paediatric cancer patients, 2 years after diagnosis. Psychooncology 2004;13(8):499-511
175. Hawthorne G, Osborne R. Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. AustNZJPublic Health 2005;29(2):136-42
176. Green DM, Whitton JA, Stovall M, et al. Pregnancy outcome of female survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. AmJ ObstetGynecol 2002;187(4):1070-80
177. Reulen RC, Zeegers MP, Wallace WH, et al. Pregnancy outcomes among adult survivors of childhood cancer in the British Childhood Cancer Survivor Study. Cancer EpidemiolBiomarkers Prev 2009;18(8):2239-47
178. Signorello LBM. Stillbirth and neonatal death in relation to radiation exposure before conception: A retrospective cohort study. The Lancet 2010(9741):624-30
179. Olson JE, Shu XO, Ross JA, et al. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. AmJ Epidemiol 1997;145(1):58-67
180. Rice F, Lewis A, Harold G, et al. Agreement between maternal report and antenatal records for a range of pre and peri-natal factors: the influence of maternal and child characteristics. Early HumDev 2007;83(8):497-504
181. Coolman M, de Groot CJ, Jaddoe VW, et al. Medical record validation of maternally reported history of preeclampsia. J Clin Epidemiol 2010;63(8):932-37
182. Cooney MA, Buck Louis GM, Sundaram R, et al. Validity of self-reported time to pregnancy. Epidemiology 2009;20(1):56-59
183. Overbeek A, van den Berg MH, Kremer LC, et al. A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges. BMCCancer 2012;12(1):363
184. de Boer EJ, den Tonkelaar I, te Velde ER, et al. Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment. HumReprod 2003;18(7):1544-52
185. van Leeuwen FE, Klip H, Mooij TM, et al. Risk of borderline and invasive ovarian tumours after ovarian stimulation for in vitro fertilization in a large Dutch cohort. HumReprod 2011;26(12):3456-65
186. The Netherlands Perinatal Registry / Stichting Perinatale Registratie N. Perinatal Care in the Netherlands 2008. Utrecht: Stichting Perinatale Registratie Nederland, 2011, 2011.

187. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74
188. Schwartz LA, Mao JJ, Derosa BW, et al. Self-reported health problems of young adults in clinical settings: survivors of childhood cancer and healthy controls. J AmBoard FamMed 2010;23(3):306-14
189. Taylor N, Absolom K, Michel G, et al. Comparison of self-reported late effects with medical records among survivors of childhood cancer. EurJ Cancer 2010;46(6):1069-78
190. O'Leary TE, Diller L, Recklitis CJ. The effects of response bias on self-reported quality of life among childhood cancer survivors. QualLife Res 2007;16(7):1211-20
191. Sou SC, Chen WJ, Hsieh WS, et al. Severe obstetric complications and birth characteristics in preterm or term delivery were accurately recalled by mothers. JClinEpidemiol 2006;59(4):429-35
192. Tomeo CA, Rich-Edwards JW, Michels KB, et al. Reproducibility and validity of maternal recall of pregnancy-related events. Epidemiology 1999;10(6):774-77
193. Union IT. The world in 2010. ICT facts and figures. . Secondary The world in 2010. ICT facts and figures. 2010. http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf
194. Group MM. Internet World Stats. Secondary Internet World Stats. 2010. http://www.internetworldstats.com.
195. Hoonakker P, Carayon P. Questionnaire Survey Nonresponse: A Comparison of Postal Mail and Internet Surveys. International Journal of Human-Computer Interaction 2009;25(5):348-73 doi: 10.1080/10447310902864951[published Online First: Epub Date]|.
196. Pealer LN, Weiler RM, Pigg RM, Jr., et al. The feasibility of a web-based surveillance system to collect health risk behavior data from college students. Health education & behavior : the official publication of the Society for Public Health Education 2001;28(5):547-59
197. Vereecken CA, Maes L. Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. The Journal of adolescent health : official publication of the Society for Adolescent Medicine 2006;38(4):426-32 doi: 10.1016/j.jadohealth.2004.10.010[published Online First: Epub Date]|.
198. Klovning A, Sandvik H, Hunskaar S. Web-based survey attracted age-biased sample with more severe illness than paper-based survey. J Clin Epidemiol 2009;62(10):1068-74 doi: 10.1016/j.jclinepi.2008.10.015[published Online First: Epub Date]|.
199. Braithwaite D, Emery J, De Lusignan S, et al. Using the Internet to conduct surveys of health professionals: a valid alternative? Family practice 2003;20(5):545-51
200. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002;4(2):E13 doi: 10.2196/jmir.4.2.e13[published Online First: Epub Date]|.
201. Fleming CM, Bowden M. Web-based surveys as an alternative to traditional mail methods. Journal of environmental management 2009;90(1):284-92 doi: 10.1016/j.jenvman.2007.09.011[published Online First: Epub Date]|.
202. Dillman DA, Smyth JD. Design effects in the transition to web-based surveys. American journal of preventive medicine 2007;32(5 Suppl):S90-6 doi: 10.1016/j.amepre.2007.03.008[published Online First: Epub Date]|.
203. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. Evaluation review 2009;33(5):464-80 doi: 10.1177/0193841x09340214[published Online First: Epub Date]|.
204. Sax LJG, S. K. ; Bryant, A. N. Assessing response rates and nonresponse bias in Web and paper surveys. Research in Higher Education 2003;44(4):409-32
205. Kiernan NE, Kiernan M, Oyler MA, et al. Is a Web Survey as Effective as a Mail Survey? A Field Experiment Among Computer Users. American Journal of Evaluation 2005;26(2):245-52 doi: 10.1177/1098214005275826[published Online First: Epub Date]|.

206. McCabe SE. Comparison of web and mail surveys in collecting illicit drug use data: a randomized experiment. Journal of drug education 2004;34(1):61-72
207. Guise V, Chambers M, Valimaki M, et al. A mixed-mode approach to data collection: combining web and paper questionnaires to examine nurses' attitudes to mental illness. Journal of advanced nursing 2010;66(7):1623-32 doi: 10.1111/j.1365-2648.2010.05357.x[published Online First: Epub Date]|.
208. McCabe SE, Diez A, Boyd CJ, et al. Comparing web and mail responses in a mixed mode survey in college alcohol use research. Addictive behaviors 2006;31(9):1619-27 doi: 10.1016/j.addbeh.2005.12.009[published Online First: Epub Date]|.
209. Chizawsky LL, Estabrooks CA, Sales AE. The feasibility of Web-based surveys as a data collection tool: a process evaluation. Applied nursing research : ANR 2011;24(1):37-44 doi: 10.1016/j.apnr.2009.03.006[published Online First: Epub Date]|.
210. Hawkins MM, Robison LL. Importance of clinical and epidemiological research in defining the long-term clinical care of pediatric cancer survivors. Pediatr Blood Cancer 2006;46(2):174-8 doi: 10.1002/pbc.20609[published Online First: Epub Date]|.
211. Hawkins MM, Lancashire ER, Winter DL, et al. The British Childhood Cancer Survivor Study: Objectives, methods, population structure, response rates and initial descriptive information. Pediatr Blood Cancer 2008;50(5):1018-25 doi: 10.1002/pbc.21335[published Online First: Epub Date]|.
212. Robison LL, Mertens AC, Boice JD, et al. Study design and cohort characteristics of the Childhood Cancer Survivor Study: a multi-institutional collaborative project. Medical and pediatric oncology 2002;38(4):229-39
213. Thompson AL, Marsland AL, Marshal MP, et al. Romantic relationships of emerging adult survivors of childhood cancer. Psycho-oncology 2009;18(7):767-74 doi: 10.1002/pon.1471[published Online First: Epub Date]|.
214. Cantrell MA, Lupinacci P. Investigating the determinants of health-related quality of life among childhood cancer survivors. Journal of advanced nursing 2008;64(1):73-83 doi: 10.1111/j.1365-2648.2008.04760.x[published Online First: Epub Date]|.
215. Finnegan L, Wilkie DJ, Wilbur J, et al. Correlates of physical activity in young adult survivors of childhood cancers. Oncology nursing forum 2007;34(5):E60-9 doi: 10.1188/07.onf.e60-e69[published Online First: Epub Date]|.
216. de Boer EJ, den Tonkelaar I, te Velde ER, et al. Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment. Human reproduction (Oxford, England) 2003;18(7):1544-52
217. Internet versus paper survey administration: Preliminary finding on response rates. 42nd Annual Conference of the International Military Testing Association; 2000; Edinburgh, Scotland.
218. Twisk JWR. Inleiding in de toegepaste biostatistiek [Introduction to applied biostatistics]. Maarssen: Elsevier Gezondheidszorg, 2007.
219. Partridge AH, Gelber S, Peppercorn J, et al. Fertility and menopausal outcomes in young breast cancer survivors. Clinical breast cancer 2008;8(1):65-9 doi: 10.3816/CBC.2008.n.004[published Online First: Epub Date]|.
220. Partridge AH, Gelber S, Peppercorn J, et al. Web-based survey of fertility issues in young women with breast cancer. J Clin Oncol 2004;22(20):4174-83 doi: 10.1200/jco.2004.01.159[published Online First: Epub Date]|.
221. Shrestha A, Nohr EA, Bech BH, et al. Parental age at childbirth and age of menarche in the offspring. Hum Reprod 2010;25(3):799-804 doi: 10.1093/humrep/dep473[published Online First: Epub Date]|.
222. Shih T-H, Xitao F. Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. Field Methods 2008;20(3):249-71 doi: 10.1177/1525822x08317085[published Online First: Epub Date]|.

223. Fan W, Yan Z. Factors affecting response rates of the web survey: A systematic review. Computers in Human Behavior 2010;26(2):132-39 doi: http://dx.doi.org/10.1016/j.chb.2009.10.015%5Bpublished Online First: Epub Date]|.
224. Langeveld NE, Grootenhuis MA, Voute PA, et al. Quality of life, self-esteem and worries in young adult survivors of childhood cancer. Psycho-oncology 2004;13(12):867-81 doi: 10.1002/pon.800[published Online First: Epub Date]|.
225. Zebrack BJ, Casillas J, Nohr L, et al. Fertility issues for young adult survivors of childhood cancer. Psycho-oncology 2004;13(10):689-99 doi: 10.1002/pon.784[published Online First: Epub Date]|.
226. Mavis BE, Brocato JJ. Postal surveys versus electronic mail surveys. The tortoise and the hare revisited. Evaluation & the health professions 1998;21(3):395-408
227. Schonlau M, Asch BJ, Du C. Web Surveys as Part of a Mixed-Mode Strategy for Populations that cannot be Contacted by E-Mail. Social Science Computer Review 2003;21(2):218-22 doi: 10.1177/0894439303021002007[published Online First: Epub Date]|.
228. Smith B, Smith TC, Gray GC, et al. When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. Am J Epidemiol 2007;166(11):1345-54 doi: 10.1093/aje/kwm212[published Online First: Epub Date]|.
229. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. Bmj 2002;324(7347):1183
230. Galesic M, Bosnjak M. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. Public Opinion Quarterly 2009;73(2):349-60 doi: 10.1093/poq/nfp031[published Online First: Epub Date]|.
231. Walston JT, Lissitz RW, Rudner LM. The influence of web-based questionnaire presentation variations on survey cooperation and perceptions of survey quality. Journal of Official Statistics 2006;22(2):271-91
232. Eaton DK, Brener ND, Kann L, et al. Comparison of paper-and-pencil versus Web administration of the Youth Risk Behavior Survey (YRBS): risk behavior prevalence estimates. Evaluation review 2010;34(2):137-53 doi: 10.1177/0193841x10362491[published Online First: Epub Date]|.
233. Connelly NA, Brown TL, Decker DJ. Factors Affecting Response Rates to Natural Resource - Focused Mail Surveys: Empirical Evidence of Declining Rates Over Time. Society & Natural Resources 2003;16(6):541-49 doi: 10.1080/08941920309152[published Online First: Epub Date]|.
234. De Leeuw E, De Heer W. Trends in household survey nonresponse: A longitudinal and international comparison. In: Groves RM, Dillman DA, Eltinge JL, et al., eds. Survey Nonresponse. New York, NY, USA: John Wiley & Sons, Inc, 2002:41-54.
235. Leisenring WM, Mertens AC, Armstrong GT, et al. Pediatric cancer survivorship research: experience of the Childhood Cancer Survivor Study. J Clin Oncol 2009;27(14):2319-27 doi: 10.1200/jco.2008.21.1813[published Online First: Epub Date]|.
236. Lee MM, Donahoe PK. Mullerian inhibiting substance: a gonadal hormone with multiple functions. EndocrRev 1993;14(2):152-64
237. Bersinger NA, Wunder D, Birkhauser MH, et al. Measurement of anti-mullerian hormone by Beckman Coulter ELISA and DSL ELISA in assisted reproduction: differences between serum and follicular fluid. Clin ChimActa 2007;384(1-2):174-75
238. Freour T, Mirallie S, Bach-Ngohou K, et al. Measurement of serum anti-Mullerian hormone by Beckman Coulter ELISA and DSL ELISA: comparison and relevance in assisted reproduction technology (ART). Clin ChimActa 2007;375(1-2):162-64
239. Hehenkamp WJ, Looman CW, Themmen AP, et al. Anti-Mullerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. JClinEndocrinolMetab 2006;91(10):4057-63

240. Streuli I, Fraisse T, Chapron C, et al. Clinical uses of anti-Mullerian hormone assays: pitfalls and promises. FertilSteril 2008
241. Kumar A, Kalra B, Patel A, et al. Development of a second generation anti-M¸llerian hormone (AMH) ELISA. Journal of Immunological Methods 2010;362(1-2):51-59 doi: 10.1016/j.jim.2010.08.011[published Online First: Epub Date]|.
242. Lee MM, Donahoe PK, Hasegawa T, et al. Mullerian inhibiting substance in humans: normal levels from infancy to adulthood. JClinEndocrinolMetab 1996;81(2):571-76
243. Seifer DB, Golub ET, Lambert-Messerlian G, et al. Variations in serum mullerian inhibiting substance between white, black, and Hispanic women. FertilSteril 2008
244. Freeman EW, Gracia CR, Sammel MD, et al. Association of anti-mullerian hormone levels with obesity in late reproductive-age women. FertilSteril 2007;87(1):101-06
245. Park AS, Lawson MA, Chuan SS, et al. Serum anti-mullerian hormone concentrations are elevated in oligomenorrheic girls without evidence of hyperandrogenism. J Clin Endocrinol Metab 2010;95(4):1786-92
246. Su HI, Sammel MD, Freeman EW, et al. Body size affects measures of ovarian reserve in late reproductive age women. Menopause 2008;15(5):857-61
247. Halawaty S, ElKattan E, Azab H, et al. Effect of obesity on parameters of ovarian reserve in premenopausal women. J ObstetGynaecolCan 2010;32(7):687-90
248. Sowers MR, McConnell D, Yosef M, et al. Relating smoking, obesity, insulin resistance, and ovarian biomarker changes to the final menstrual period. AnnNY AcadSci 2010;1204:95-103
249. Appt SE, Clarkson TB, Chen H, et al. Serum antimullerian hormone predicts ovarian reserve in a monkey model. Menopause 2009;16(3):597-601
250. Cook CL, Siow Y, Taylor S, et al. Serum mullerian-inhibiting substance levels during normal menstrual cycles. FertilSteril 2000;73(4):859-61
251. Lahlou N, Chabbert-Buffet N, Gainer E, et al. O-24: Diphasic pattern of anti-mullerian hormone (AMH) in ovulatory cycles as evidenced by means of an ultra-sensitive assay: New insights into ovarian function. Fertility and Sterility 2006;86(3, Supplement 1):S11-S11 doi: 10.1016/j.fertnstert.2006.07.030[published Online First: Epub Date]|.
252. Rustamov O, Pemberton PW, Roberts SA, et al. The reproducibility of serum anti-Mullerian hormone in subfertile women: within and between patient variability. FertilSteril 2010
253. Wunder DM, Bersinger NA, Yared M, et al. Statistically significant changes of antimullerian hormone and inhibin levels during the physiologic menstrual cycle in reproductive age women. FertilSteril 2008;89(4):927-33
254. Sowers M, McConnell D, Gast K, et al. Anti-Mullerian hormone and inhibin B variability during normal menstrual cycles. FertilSteril 2010;94(4):1482-86
255. Roberts SA. Variability in anti-Mullerian hormone levels: a comment on Sowers et al., Anti-Mullerian hormone and inhibin B variability during normal menstrual cycles. FertilSteril 2010;94(4):e59
256. Broer SL, Mol B, Dolleman M, et al. The role of anti-Mullerian hormone assessment in assisted reproductive technology outcome. CurrOpinObstetGynecol 2010;22(3):193-201
257. van Disseldorp J, Lambalk CB, Kwee J, et al. Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. HumReprod 2010;25(1):221-27
258. Nelson SM, Messow MC, Wallace AM, et al. Nomogram for the decline in serum antimullerian hormone: a population study of 9,601 infertility patients. FertilSteril 2011;95(2):736-41
259. van Disseldorp J, Faddy MJ, Themmen AP, et al. Relationship of serum antimullerian hormone concentration to age at menopause. J Clin Endocrinol Metab 2008;93(6):2129-34

260. van Rooij I, Broekmans FJ, Scheffer GJ, et al. Serum antimullerian hormone levels best reflect the reproductive decline with age in normal women with proven fertility: a longitudinal study. FertilSteril 2005;83(4):979-87
261. Shebl O, Ebner T, Sir A, et al. Age-related distribution of basal serum AMH level in women of reproductive age and a presumably healthy cohort. Fertility and Sterility 2011;95(2):832-34 doi: 10.1016/j.fertnstert.2010.09.012[published Online First: Epub Date]|.
262. Sowers MR, Eyvazzadeh AD, McConnell D, et al. Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. JClinEndocrinolMetab 2008;93(9):3478-83
263. Lambalk CB, van Disseldorp J, de Koning CH, et al. Testing ovarian reserve to predict age at menopause. Maturitas 2009;63(4):280-91
264. Jayaprakasan K, Campbell BK, Clewes JS, et al. Three-dimensional ultrasound improves the interobserver reliability of antral follicle counts and facilitates increased clinical work flow. Ultrasound ObstetGynecol 2008;31(4):439-44
265. Moawad NS, Gibbons H, Liu J, et al. Comparison of 3- and 2-dimensional sonographic techniques for counting ovarian follicles. J Ultrasound Med 2009;28(10):1281-88
266. Merce LT, Gomez B, Engels V, et al. Intraobserver and interobserver reproducibility of ovarian volume, antral follicle count, and vascularity indices obtained with transvaginal 3-dimensional ultrasonography, power Doppler angiography, and the virtual organ computer-aided analysis imaging program. J Ultrasound Med 2005;24(9):1279-87
267. Martins WP, Nastri CO. Interpreting reproducibility results for ultrasound measurements. Ultrasound Obstet Gynecol 2014;43(4):479-80 doi: 10.1002/uog.13320[published Online First: Epub Date]|.
268. Scheffer GJ, Broekmans FJ, Bancsi LF, et al. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound ObstetGynecol 2002;20(3):270-75
269. Deb S, Campbell BK, Clewes JS, et al. Quantitative analysis of antral follicle number and size: a comparison of two-dimensional and automated three-dimensional ultrasound techniques. Ultrasound ObstetGynecol 2010;35(3):354-60
270. Sklar CA, Mertens AC, Mitby P, et al. Premature menopause in survivors of childhood cancer: a report from the childhood cancer survivor study. J NatlCancer Inst 2006;98(13):890-96
271. Overbeek A, van den Berg MH, Kremer LC, et al. A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges. BMCCancer 2012;12:363
272. Jayaprakasan K, Walker KF, Clewes JS, et al. The interobserver reliability of off-line antral follicle counts made from stored three-dimensional ultrasound data: a comparative study of different measurement techniques. Ultrasound ObstetGynecol 2007;29(3):335-41
273. Liao JJ. Sample size calculation for an agreement study. PharmStat 2010;9(2):125-32
274. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10
275. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45(1):255-68
276. Coelho Neto MA, Roncato P, Nastri CO, et al. True Reproducibility of UltraSound Techniques (TRUST): systematic review of reliability studies in obstetrics and gynecology. Ultrasound Obstet Gynecol 2014 doi: 10.1002/uog.14654[published Online First: Epub Date]|.
277. Deb S, Jayaprakasan K, Campbell BK, et al. Intraobserver and interobserver reliability of automated antral follicle counts made using three-dimensional ultrasound and SonoAVC. Ultrasound ObstetGynecol 2009;33(4):477-83

278. Skouby SO. Contraceptive use and behavior in the 21st century: a comprehensive study across five European countries. EurJContraceptReprodHealth Care 2004;9(2):57-68
279. Dericks-Tan JS, Gudacker V, Taubert HD. Influence of oral contraceptives on integrated secretion of gonadotropins. Contraception 1992;46(4):369-77
280. Mishell DR, Jr., Kletzky OA, Brenner PF, et al. The effect of contraceptive steroids on hypothalamic-pituitary function. AmJObstetGynecol 1977;128(1):60-74
281. van der Spuy ZM, Sohnius U, Pienaar CA, et al. Gonadotropin and estradiol secretion during the week of placebo therapy in oral contraceptive pill users. Contraception 1990;42(6):597-609
282. Vandever MA, Kuehl TJ, Sulak PJ, et al. Evaluation of pituitary-ovarian axis suppression with three oral contraceptive regimens. Contraception 2008;77(3):162-70
283. Baerwald AR, Olatunbosun OA, Pierson RA. Ovarian follicular development is initiated during the hormone-free interval of oral contraceptive use. Contraception 2004;70(5):371-77
284. Schlaff WD, Lynch AM, Hughes HD, et al. Manipulation of the pill-free interval in oral contraceptive pill users: the effect on follicular suppression. AmJObstetGynecol 2004;190(4):943-51
285. van Heusden AM, Fauser BC. Activity of the pituitary-ovarian axis in the pill-free interval during use of low-dose combined oral contraceptives. Contraception 1999;59(4):237-43
286. Willis SA, Kuehl TJ, Spiekerman AM, et al. Greater inhibition of the pituitary--ovarian axis in oral contraceptive regimens with a shortened hormone-free interval. Contraception 2006;74(2):100-03
287. van Heusden AM, Fauser BC. Residual ovarian activity during oral steroid contraception. HumReprodUpdate 2002;8(4):345-58
288. van Heusden AM, Coelingh Bennink HJ, Fauser BC. FSH and ovarian response: spontaneous recovery of pituitary-ovarian activity during the pill-free period vs. exogenous recombinant FSH during high-dose combined oral contraceptives. ClinEndocrinol(Oxf) 2002;56(4):509-17
289. Cohen BL, Katz M. Pituitary and ovarian function in women receiving hormonal contraception. Contraception 1979;20(5):475-87
290. Fauser BC, van Heusden AM. Manipulation of human ovarian function: physiological concepts and clinical consequences. EndocrRev 1997;18(1):71-106
291. Somunkiran A, Yavuz T, Yucel O, et al. Anti-Mullerian hormone levels during hormonal contraception in women with polycystic ovary syndrome. EurJObstetGynecolReprodBiol 2007;134(2):196-201
292. de Graaf A. Geboorteregeling in 2008 [Birth regulation in 2008]. In: Bevolkingstrends [Population trends] Centraal Bureau voor de Statistiek [Statistics Netherlands] pp. 54-59, 2009.
293. de Boer EJ, den TI, te Velde ER, et al. Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment. HumReprod 2003;18(7):1544-52
294. de Vet A, Laven JS, de Jong FH, et al. Antimullerian hormone serum levels: a putative marker for ovarian aging. FertilSteril 2002;77(2):357-62
295. La Marca A, Broekmans FJ, Volpe A, et al. Anti-Mullerian hormone (AMH): what do we still need to know? HumReprod 2009;24(9):2264-75
296. Rosen A, Rodriguez-Wallberg KA, Rosenzweig L. Psychosocial distress in young cancer survivors. SeminOncolNurs 2009;25(4):268-77
297. Zebrack B. Information and service needs for young adult cancer survivors. SupportCare Cancer 2009;17(4):349-57
298. Zebrack BJ, Mills J, Weitzman TS. Health and supportive care needs of young adult cancer patients and survivors. JCancer Surviv 2007;1(2):137-45
299. Jernstrom H, Knutsson M, Olsson H. Temporary increase of FSH levels in healthy, nulliparous, young women after cessation of low-dose oral contraceptive use. Contraception 1995;52(1):51-56
300. Lambalk CB, Schoemaker J. Hypothetical risks of twinning in the natural menstrual cycle. EurJObstetGynecolReprodBiol 1997;75(1):1-4

301. Renier MA, Vereecken A, Van HE, et al. Dimeric inhibin serum values as markers of ovarian activity in pill-free intervals. Contraception 1998;57(1):45-48
302. Streuli I, Fraisse T, Pillet C, et al. Serum antimullerian hormone levels remain stable throughout the menstrual cycle and after oral or vaginal administration of synthetic sex steroids. FertilSteril 2008;90(2):395-400
303. Broekmans FJ, Visser JA, Laven JS, et al. Anti-Mullerian hormone and ovarian dysfunction. Trends EndocrinolMetab 2008;19(9):340-47
304. Durlinger AL, Visser JA, Themmen AP. Regulation of ovarian function: the role of anti-Mullerian hormone. Reproduction 2002;124(5):601-09
305. Fanchin R, Taieb J, Lozano DH, et al. High reproducibility of serum anti-Mullerian hormone measurements suggests a multi-staged follicular secretion and strengthens its role in the assessment of ovarian follicular status. HumReprod 2005;20(4):923-27
306. La Marca A, Malmusi S, Giulini S, et al. Anti-Mullerian hormone plasma levels in spontaneous menstrual cycle and during treatment with FSH to induce ovulation. HumReprod 2004;19(12):2738-41
307. Streuli I, Fraisse T, Chapron C, et al. Clinical uses of anti-Mullerian hormone assays: pitfalls and promises. FertilSteril 2009;91(1):226-30
308. van Disseldorp J, Lambalk CB, Kwee J, et al. Comparison of inter- and intra-cycle variability of Anti-Mullerian Hormone and antral follicle counts. HumReprod 2009
309. Elter K, Sismanoglu A, Durmusoglu F. Intercycle variabilities of basal antral follicle count and ovarian volume in subfertile women and their relationship to reproductive aging: a prospective study. GynecolEndocrinol 2005;20(3):137-43
310. Kwee J, Schats R, McDonnell J, et al. Intercycle variability of ovarian reserve tests: results of a prospective randomized study. HumReprod 2004;19(3):590-95
311. Lenton EA, Lawrence GF, Coleman RA, et al. Individual variation in gonadotrophin and steroid concentrations and in the lengths of the follicular and luteal phases in women with regular menstrual cycles. ClinReprodFertil 1983;2(2):143-50
312. Penarrubia J, Fabregues F, Manau D, et al. Initial analysis of variability among basal hormone biomarkers of ovarian reserve. ReprodBiomedOnline 2004;8(2):191-95
313. Scheffer GJ, Broekmans FJ, Dorland M, et al. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. FertilSteril 1999;72(5):845-51
314. Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. EndocrRev 2009;30(5):465-93
315. Gorman JR, Su HI, Roberts SC, et al. Experiencing reproductive concerns as a female cancer survivor is associated with depression. Cancer 2015;121(6):935-42 doi: 10.1002/cncr.29133[published Online First: Epub Date]|.
316. Barton SE, Najita JS, Ginsburg ES, et al. Infertility, infertility treatment, and achievement of pregnancy in female survivors of childhood cancer: a report from the Childhood Cancer Survivor Study cohort. Lancet Oncol 2013;14(9):873-81 doi: 10.1016/S1470-2045(13)70251-1[published Online First: Epub Date]|.
317. Su HI, Sammel MD, Green J, et al. Antimullerian hormone and inhibin B are hormone measures of ovarian function in late reproductive-aged breast cancer survivors. Cancer 2010;116(3):592-99
318. Dewailly D, Andersen CY, Balen A, et al. The physiology and clinical utility of anti-Mullerian hormone in women. Hum Reprod Update 2014;20(3):370-85 doi: 10.1093/humupd/dmt062[published Online First: Epub Date]|.
319. Iliodromiti S, Anderson RA, Nelson SM. Technical and performance characteristics of anti-Mullerian hormone and antral follicle count as biomarkers of ovarian response. Hum Reprod Update 2015;21(6):698-710 doi: 10.1093/humupd/dmu062[published Online First: Epub Date]|.

320. Overbeek A, van den Berg MH, van Leeuwen FE, et al. Chemotherapy-related late adverse effects on ovarian function in female survivors of childhood and young adult cancer: A systematic review. Cancer treatment reviews 2017;53:10-24 doi: 10.1016/j.ctrv.2016.11.006[published Online First: Epub Date]|.
321. Green DM, Kawashima T, Stovall M, et al. Fertility of female survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. J Clin Oncol 2009;27:2677-85
322. Bozdag G, Calis P, Zengin D, et al. Age related normogram for antral follicle count in general population and comparison with previous studies. European journal of obstetrics, gynecology, and reproductive biology 2016;206:120-24 doi: 10.1016/j.ejogrb.2016.09.013[published Online First: Epub Date]|.
323. Hall JE. Endocrinology of the Menopause. Endocrinology and metabolism clinics of North America 2015;44(3):485-96 doi: 10.1016/j.ecl.2015.05.010[published Online First: Epub Date]|.
324. Metzger ML, Meacham LR, Patterson B, et al. Female reproductive health after childhood, adolescent, and young adult cancers: guidelines for the assessment and management of female reproductive complications. J Clin Oncol 2013;31:1239-47
325. van Dorp WM, R.L.; Kremer, L.C.; Hudson, M.M.; van den Heuvel-Eibrink, M.M.; van den Berg, M.H.; Levine J.M.; van Dulmen-den Broeder, E.; di Iorgi, N.; Albanese, A.; Armenian, S.H.; Bhatia, S.; Constine, S.L.; Corrias, A.; Deans, R.; Dirksen, U.; Gracia, C.; Hjorth, L.; Kroon, L.; Lambalk, C.B.; Landier, W.; Levitt, G.; Leiper, A.; Meacham, L.; Mussa, A.; Neggers, S.J.; Oeffinger, K.C.; Revelli, A.; van Santen, H.M.; Skinner, R.; Toogood, A.; Wallace, W.H.; Haupt, R. Recommendations for Premature Ovarian Insufficiency Surveillance for Female Childhood, Adolescent and Young Adult Cancer Survivors: A Report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. Accepted for publication in Journal of Clinical Oncology (March 8, 2016)
326. Green DM, Kawashima T, Stovall M, et al. Fertility of female survivors of childhood cancer: a report from the childhood cancer survivor study. J Clin Oncol 2009;27(16):2677-85 doi: JCO.2008.20.1541 [pii];10.1200/JCO.2008.20.1541 [doi][published Online First: Epub Date]|.
327. Tucker MA, Meadows AT, Boice JD, Jr., et al. Leukemia after therapy with alkylating agents for childhood cancer. J Natl Cancer Inst 1987;78(3):459-64
328. Green DM, Nolan VG, Goodman PJ, et al. The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: a report from the Childhood Cancer Survivor Study. Pediatr Blood Cancer 2014;61(1):53-67 doi: 10.1002/pbc.24679 [doi][published Online First: Epub Date]|.
329. Smith MA, Seibel NL, Altekruse SF, et al. Outcomes for children and adolescents with cancer: challenges for the twenty-first century. JClinOncol 2010;28(15):2625-34
330. Byrne JF. Early menopause in long-term survivors of cancer during adolescence. American Journal of Obstetrics and Gynecology 1992(3):788-93
331. Chemaitilly W, Mertens AC, Mitby P, et al. Acute ovarian failure in the childhood cancer survivor study. The Journal of clinical endocrinology and metabolism 2006;91(5):1723-28
332. Chiarelli AM, Marrett LD, Darlington G. Early menopause and infertility in females after treatment for childhood cancer diagnosed in 1964-1988 in Ontario, Canada. AmJ Epidemiol 1999;150(3):245-54
333. Mueller BA, Chow EJ, Kamineni A, et al. Pregnancy outcomes in female childhood and adolescent cancer survivors: a linked cancer-birth registry analysis. ArchPediatrAdolescMed 2009;163(10):879-86
334. Winther JFB. Spontaneous abortion in a Danish population-based cohort of childhood cancer survivors. Journal of Clinical Oncology 2008(26):4340-46
335. Lantinga GM, Simons AH, Kamps WA, et al. Imminent ovarian failure in childhood cancer survivors. EurJ Cancer 2006;42(10):1415-20
336. Larsen EC, Muller J, Rechnitzer C, et al. Diminished ovarian reserve in female childhood cancer survivors with regular menstrual cycles and basal FSH <10 IU/l. HumReprod 2003;18(2):417-22

337. Lie Fong S, Van Den Heuvel-Eibrink M. Pregnancy outcome in female childhood cancer survivors. Molecular Human ReproductionConference: 25th Annual Meeting of the European Society of Human Reproduction and Embryology, ESHRE Amsterdam NetherlandsConference Start: 20090628 Conference End: 20090701Conference Publication 2009(var.pagings):i101
338. Arafa MA, Rabah DM. Attitudes and practices of oncologists toward fertility preservation. JPediatrHematolOncol 2011;33(3):203-07
339. Collins IM, Fay L, Kennedy MJ. Strategies for fertility preservation after chemotherapy: awareness among Irish cancer specialists. IrMedJ 2011;104(1):6-9
340. Duffy C, Allen SM, Dube C, et al. Oncologists' confidence in knowledge of fertility issues for young women with cancer. JCancer Educ 2012;27(2):369-76
341. Forman EJ, Anders CK, Behera MA. A nationwide survey of oncologists regarding treatment-related infertility and fertility preservation in female cancer patients. FertilSteril 2010;94(5):1652-56
342. Quinn GP, Vadaparampil ST. Fertility preservation and adolescent/young adult cancer patients: physician communication challenges. JAdolescHealth 2009;44(4):394-400
343. Rabah DM, El-Nimr N, Rafe BA, et al. Fertility cryopreservation for female cancer patients: attitudes and clinical practices of oncologists in Riyadh, Saudi Arabia. JReprodMed 2012;57(9-10):431-34
344. Yee S, Fuller-Thomson E, Lau A, et al. Fertility preservation practices among Ontario oncologists. JCancer Educ 2012;27(2):362-68
345. Anderson RA, Weddell A, Spoudeas HA, et al. Do doctors discuss fertility issues before they treat young patients with cancer? HumReprod 2008;23(10):2246-51
346. Goodwin T, Elizabeth OB, Kiernan M, et al. Attitudes and practices of pediatric oncology providers regarding fertility issues. PediatrBlood Cancer 2007;48(1):80-85
347. Kohler TS, Kondapalli LA, Shah A, et al. Results from the survey for preservation of adolescent reproduction (SPARE) study: gender disparity in delivery of fertility preservation message to adolescents with cancer. JAssistReprodGenet 2011;28(3):269-77
348. Detti L, Martin DC, Williams LJ. Applicability of adult techniques for ovarian preservation to childhood cancer patients. JAssistReprodGenet 2012;29(9):985-95
349. Meirow D, Nugent D. The effects of radiotherapy and chemotherapy on female reproduction. HumReprodUpdate 2001;7(6):535-43
350. Wallace WH, Kelsey TW, Anderson RA. Ovarian cryopreservation: experimental or established and a cure for the menopause? ReprodBiomedOnline 2012;25(2):93-95
351. Kissell KA, Danaher MR, Schisterman EF, et al. Biological variability in serum anti-Mullerian hormone throughout the menstrual cycle in ovulatory and sporadic anovulatory cycles in eumenorrheic women. Human reproduction (Oxford, England) 2014;29(8):1764-72 doi: 10.1093/humrep/deu142[published Online First: Epub Date]|.
352. Randolph JF, Jr., Harlow SD, Helmuth ME, et al. Updated assays for inhibin B and AMH provide evidence for regular episodic secretion of inhibin B but not AMH in the follicular phase of the normal menstrual cycle. Human reproduction (Oxford, England) 2014;29(3):592-600 doi: 10.1093/humrep/det447[published Online First: Epub Date]|.
353. Robertson DM, Hale GE, Fraser IS, et al. Changes in serum antimullerian hormone levels across the ovulatory menstrual cycle in late reproductive age. Menopause (New York, NY) 2011;18(5):521-4 doi: 10.1097/gme.0b013e3181f8d9e0[published Online First: Epub Date]|.
354. Kelsey TW, Wright P, Nelson SM, et al. A validated model of serum anti-mullerian hormone from conception to menopause. PloS one 2011;6(7):e22024 doi: 10.1371/journal.pone.0022024[published Online First: Epub Date]|.

355. La Marca A, Sighinolfi G, Giulini S, et al. Normal serum concentrations of anti-Mullerian hormone in women with regular menstrual cycles. Reprod Biomed Online 2010;21(4):463-9 doi: 10.1016/j.rbmo.2010.05.009[published Online First: Epub Date]|.
356. La Marca A, Spada E, Grisendi V, et al. Normal serum anti-Mullerian hormone levels in the general female population and the relationship with reproductive history. European journal of obstetrics, gynecology, and reproductive biology 2012;163(2):180-4 doi: 10.1016/j.ejogrb.2012.04.013[published Online First: Epub Date]|.
357. Seifer DB, Baker VL, Leader B. Age-specific serum anti-Mullerian hormone values for 17,120 women presenting to fertility centers within the United States. Fertil Steril 2011;95(2):747-50 doi: 10.1016/j.fertnstert.2010.10.011[published Online First: Epub Date]|.
358. Hansen KR, Knowlton NS, Thyer AC, et al. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. Human reproduction (Oxford, England) 2008;23(3):699-708 doi: 10.1093/humrep/dem408[published Online First: Epub Date]|.
359. Wallace WH, Kelsey TW. Human ovarian reserve from conception to the menopause. PloS one 2010;5(1):1-9
360. van Dorp W, Mulder RL, Kremer LC, et al. Recommendations for Premature Ovarian Insufficiency Surveillance for Female Survivors of Childhood, Adolescent, and Young Adult Cancer: A Report From the International Late Effects of Childhood Cancer Guideline Harmonization Group in Collaboration With the PanCareSurFup Consortium. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2016 doi: 10.1200/JCO.2015.64.3288[published Online First: Epub Date]|.
361. Kedem A, Haas J, Geva LL, et al. Ongoing pregnancy rates in women with low and extremely low AMH levels. A multivariate analysis of 769 cycles. PloS one 2013;8(12):e81629 doi: 10.1371/journal.pone.0081629[published Online First: Epub Date]|.
362. Zarek SM, Mitchell EM, Sjaarda LA, et al. Is Anti-Mullerian Hormone Associated With Fecundability? Findings From the EAGeR Trial. The Journal of clinical endocrinology and metabolism 2015;100(11):4215-21 doi: 10.1210/jc.2015-2474[published Online First: Epub Date]|.
363. Reijnders IF, Nelen WL, IntHout J, et al. The value of Anti-Mullerian hormone in low and extremely low ovarian reserve in relation to live birth after in vitro fertilization. European journal of obstetrics, gynecology, and reproductive biology 2016;200:45-50 doi: 10.1016/j.ejogrb.2016.02.007[published Online First: Epub Date]|.
364. van Dorp W, Blijdorp K, Laven Js et al. Decreased ovarian function is associated with obesity in very long-term female survivors of childhood cancer. Eur J Endocrinol 2013 2013;168:905-12
365. Broer SL, Eijkemans MJ, Scheffer GJ, et al. Anti-mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. The Journal of clinical endocrinology and metabolism 2011;96(8):2532-9 doi: 10.1210/jc.2010-2776[published Online First: Epub Date]|.
366. Freeman EW, Sammel MD, Lin H, et al. Contribution of the rate of change of antimullerian hormone in estimating time to menopause for late reproductive-age women. Fertil Steril 2012;98(5):1254-9 e1-2 doi: 10.1016/j.fertnstert.2012.07.1139[published Online First: Epub Date]|.
367. Beneventi F, Locatelli E, Giorgiani G, et al. Gonadal and uterine function in female survivors treated by chemotherapy, radiotherapy, and/or bone marrow transplantation for childhood malignant and non-malignant diseases. BJOG 2014;121(7):856-65; discussion 65 doi: 10.1111/1471-0528.12715[published Online First: Epub Date]|.
368. Morse H, Elfving M, Lindgren A, et al. Acute onset of ovarian dysfunction in young females after start of cancer treatment. Pediatr Blood Cancer 2013;60:676-81

369. Andersen AN, Witjes H, Gordon K, et al. Predictive factors of ovarian response and clinical outcome after IVF/ICSI following a rFSH/GnRH antagonist protocol with or without oral contraceptive pre-treatment. Human reproduction (Oxford, England) 2011;26(12):3413-23 doi: 10.1093/humrep/der318[published Online First: Epub Date]|.
370. Deb S, Campbell BK, Pincott-Allen C, et al. Quantifying effect of combined oral contraceptive pill on functional ovarian reserve as measured by serum anti-Mullerian hormone and small antral follicle count using three-dimensional ultrasound. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 2012;39(5):574-80 doi: 10.1002/uog.10114[published Online First: Epub Date]|.
371. Li HW, Wong CY, Yeung WS, et al. Serum anti-mullerian hormone level is not altered in women using hormonal contraceptives. Contraception 2011;83(6):582-5 doi: 10.1016/j.contraception.2010.09.007[published Online First: Epub Date]|.
372. Bentzen JG, Forman JL, Pinborg A, et al. Ovarian reserve parameters: a comparison between users and non-users of hormonal contraception. Reprod Biomed Online 2012;25(6):612-9 doi: 10.1016/j.rbmo.2012.09.001[published Online First: Epub Date]|.
373. Kallio S, Puurunen J, Ruokonen A, et al. Antimullerian hormone levels decrease in women using combined contraception independently of administration route. Fertil Steril 2013;99(5):1305-10 doi: 10.1016/j.fertnstert.2012.11.034[published Online First: Epub Date]|.
374. Kristensen SL, Ramlau-Hansen CH, Andersen CY, et al. The association between circulating levels of antimullerian hormone and follicle number, androgens, and menstrual cycle characteristics in young women. Fertil Steril 2012;97(3):779-85 doi: 10.1016/j.fertnstert.2011.12.017[published Online First: Epub Date]|.
375. Lawrenz B, Fehm T, von Wolff M, et al. Reduced pretreatment ovarian reserve in premenopausal female patients with Hodgkin lymphoma or non-Hodgkin-lymphoma--evaluation by using antimullerian hormone and retrieved oocytes. Fertil Steril 2012;98(1):141-4 doi: 10.1016/j.fertnstert.2012.04.021[published Online First: Epub Date]|.
376. van Dorp W, van den Heuvel-Eibrink MM, de Vries AC, et al. Decreased serum anti-Mullerian hormone levels in girls with newly diagnosed cancer. Human reproduction (Oxford, England) 2014;29(2):337-42 doi: 10.1093/humrep/det442[published Online First: Epub Date]|.
377. van Dorp W, van den Heuvel-Eibrink MM, Stolk L, et al. Genetic variation may modify ovarian reserve in female childhood cancer survivors. Hum Reprod 2013;28(4):1069-76 doi: 10.1093/humrep/des472[published Online First: Epub Date]|.
378. Dolleman M, Faddy MJ, van Disseldorp J, et al. The relationship between anti-Mullerian hormone in women receiving fertility assessments and age at menopause in subfertile women: evidence from large population studies. The Journal of clinical endocrinology and metabolism 2013;98(5):1946-53 doi: 10.1210/jc.2012-422810.1210/jc.2013-3105[published Online First: Epub Date]|.
379. Dolleman M, Verschuren WM, Eijkemans MJ, et al. Reproductive and lifestyle determinants of anti-Mullerian hormone in a large population-based study. The Journal of clinical endocrinology and metabolism 2013;98(5):2106-15 doi: 10.1210/jc.2012-3995[published Online First: Epub Date]|.
380. Caserta D, Bordi G, Di Segni N, et al. The influence of cigarette smoking on a population of infertile men and women. Archives of gynecology and obstetrics 2013;287(4):813-8 doi: 10.1007/s00404-012-2643-5[published Online First: Epub Date]|.
381. Hayatbakhsh MR, Clavarino A, Williams GM, et al. Cigarette smoking and age of menopause: a large prospective study. Maturitas 2012;72(4):346-52 doi: 10.1016/j.maturitas.2012.05.004[published Online First: Epub Date]|.

382. Pelosi E, Simonsick E, Forabosco A, et al. Dynamics of the ovarian reserve and impact of genetic and epidemiological factors on age of menopause. Biology of reproduction 2015;92(5):130 doi: 10.1095/biolreprod.114.127381[published Online First: Epub Date]|.
383. van Dorp W, Blijdorp K, Laven JS, et al. Decreased ovarian function is associated with obesity in very long-term female survivors of childhood cancer. Eur J Endocrinol 2013;168(6):905-12 doi: 10.1530/EJE-13-0114[published Online First: Epub Date]|.
384. van Dorp W, Owusuaa C, Laven JS, et al. Characteristics and outcome of pediatric non-Hodgkin lymphoma patients with ovarian infiltration at presentation. Pediatric blood & cancer 2013;60(12):2054-9 doi: 10.1002/pbc.24582[published Online First: Epub Date]|.
385. Lass A. The fertility potential of women with a single ovary. Human reproduction update 1999;5(5):546-50
386. Dawber TR, Kannel WB. An epidemiologic study of heart disease: the Framingham study. Nutrition reviews 1958;16(1):1-4
387. Prentice RLA, G.; Cummings, S.; Freedman, L.S.; Furberg, C.; Henderson, M.; Johnson, S.R.; Kuller, L.; Manson, J.; Oberman, A.; Rossouw, J.E. . Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Controlled clinical trials 1998;19(1):61-109
388. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97
389. Hunault CC, Habbema JD, Eijkemans MJ, et al. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. Hum Reprod 2004;19(9):2019-26 doi: 10.1093/humrep/deh365[published Online First: Epub Date]|.
390. Elwood JM. Critical appraisal of epidemiological studies and clinical trials. 2007. Oxford University Press, New York.
391. van den Berg MH, van Dulmen-den Broeder E, Overbeek A, et al. Fertility studies in female childhood cancer survivors: selecting appropriate comparison groups. Reprod Biomed Online. 2014 Sep;29(3):352-61.